INTERNATIONAL JOURNAL OF NETWORK MANAGEMENT

Int. J. Network Mgmt 2018; **00**:1–22

Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/nem

Detection and Mitigation of Monitor Identification Attacks in Collaborative Intrusion Detection Systems

Emmanouil Vasilomanolakis*, Max Mühlhäuser

Telecooperation Group, Technische Universität Darmstadt, Darmstadt, Germany.

SUMMARY

Collaborative defensive approaches such as Collaborative Intrusion Detection Systems (CIDSs) have emerged as a response to the continuous increase in the sophistication of cyber-attacks. Such systems utilize a plethora of heterogeneous monitors to create a holistic picture of the monitored network. A number of research institutes deploy CIDSs that publish their alert data publicly, over the Internet. This is important for researchers and security administrators, as such systems provide a source of real-world alert data for experimentation. However, a class of identification attacks exists, namely Probe-Response Attacks (PRAs), which can significantly reduce the benefits of a CIDS. In particular, such attacks allow an adversary to detect the network location of the monitors of a CIDS. This article discusses the state of the art, with an emphasis on our previous and ongoing work, with regard to the detection and the mitigation of PRAs. We compare the most promising defensive mechanisms with respect to their effectiveness and the possible negative effects they might introduce to the CIDS. Finally, we provide a thorough discussion of research gaps and possible future directions for the field. Copyright © 2018 John Wiley & Sons, Ltd.

Received ...

KEY WORDS: probe-response attacks; collaborative intrusion detection; identification attacks

*Correspondence to: vasilomano@tk.tu-darmstadt.de, Telecooperation Group, Technische Universität Darmstadt, Darmstadt, Hochschulstr. 10, D-64289, Germany.

1. INTRODUCTION

Sophisticated and highly tailored attacks, e.g., Distributed Denial of Service (DDoS) attacks and Advanced Persistent Threats (APTs), are constantly increasing [18]. To cope with this, research in cyber-security is moving from isolated security solutions such as honeypots and Intrusion Detection Systems (IDSs) [13] towards more collaborative approaches [28, 4]. Such systems, commonly coined as Collaborative Intrusion Detection Systems (CIDSs), function by making use of a plethora of monitors, which collaborate by exchanging alert data, to create a holistic view of the monitored network [22].

Over the years a number of research institutes and corporations have deployed CIDSs which publish their alert data publicly over the Internet. For instance, the DShield [19] and TraCINg [21] CIDSs belong into this category. Figure 1, depicts a snapshot of the TraCINg system which was developed by us in our previous work [21]. In more details, a glance of such an example of publicly available alert data, in the TraCINg CIDS, is given in Figure 2. In more details, the figure depicts the way the CIDS reports alerts. For instance, in the first row one can observe a *MySQL* attack occurring from an IP address that appears to be in Indonesia (from port 16384) and is targeting a CIDS monitor in Germany (in port 3306).

These systems, also referred to as cyber incident monitors or network telescopes [15], are important for both the research community and for securing the Internet in general. For instance,

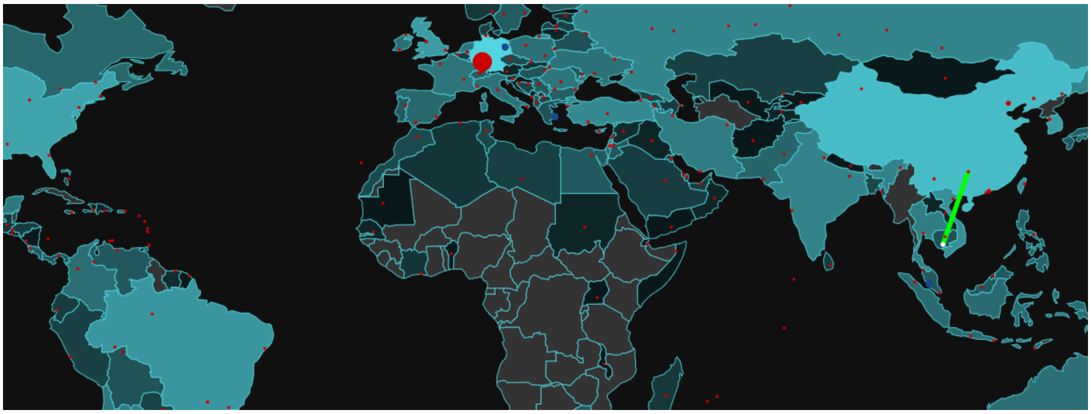
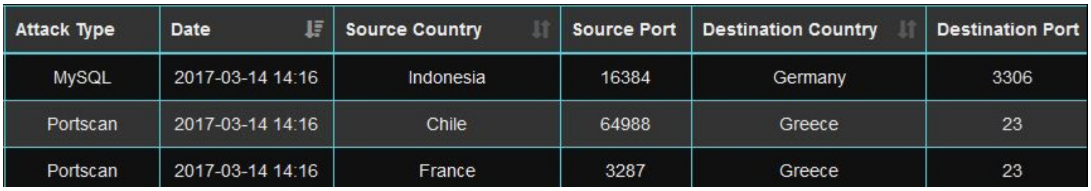


Figure 1. TraCINg CIDS alert data map overview

DOI: 10.1002/nem



Attack Type	Date	Source Country	Source Port	Destination Country	Destination Port
MySQL	2017-03-14 14:16	Indonesia	16384	Germany	3306
Portscan	2017-03-14 14:16	Chile	64988	Greece	23
Portscan	2017-03-14 14:16	France	3287	Greece	23

Figure 2. Publicly available alert data output in the TraCINg CIDS

DShield aided in the early detection of the Code-Red worm [14]. In addition, such publicly available alert data can assist researchers for their experiments. For example, alert datasets are very important for the evaluation of intrusion detection algorithms and systems [20].

A lot of research has been conducted with regard to CIDSs and potential attacks into them [9]. In particular, a class of *monitor identification* attacks exists, which makes it possible for an adversary to identify the network location, i.e., the IP addresses, of the monitors of a CIDS. These attacks are called Probe-Response Attacks (PRAs) and can have a significantly negative impact for a CIDS. For example, an attacker can utilize such knowledge to either attack the CIDS monitors, e.g., via DDoS attacks, or to create sophisticated malware that can evade monitors and thus remain undetected for a longer period of time.

In this article, we discuss the state of the art with regard to the detection and the mitigation of PRAs. In particular, the article at hand bridges and extends all our previous work [24, 25, 26, 20] and offers an overview as well as a comparison of the state of the art. We emphasize on the various proposed mitigation methods, their advantages and disadvantages, and their impact in the usability of the CIDS. Furthermore, we identify and discuss research gaps and corresponding ideas for potential future work.

At a glance, the main contributions of this article are as follows:

- The article provides an overview and introduction to the problem of PRAs.
- The article combines the work of several papers (from the view of improving, detecting and mitigating PRAs) and compares the various proposed techniques.
- Future work and research gaps have been identified for both the detection of PRAs as well as for their mitigation.

The remainder of this article is organized as follows. In Section 2, we provide background information for PRAs by thoroughly discussing how this class of attacks operates. Section 3, examines techniques for the detection of PRAs. Moreover, Section 4, discusses the methods that can be used for PRA mitigation. In addition, it compares, via a qualitative comparison, all existing mitigation techniques. Section 5, identifies research gaps and discusses ideas for future work. Finally, Section 6 concludes this article.

2. BACKGROUND

CIDSs can be classified, with respect to their network architecture, into centralized, hierarchical and distributed [22]. Each of these classes has its own advantages with regard to the scalability and the overall accuracy of the system. Nevertheless, regardless of the architecture that a CIDS is employing it is important that the monitors exchanging alert data remain *anonymous*. The reason for this is that when an adversary identifies the network location of a monitor, she can attempt to either take it down or evade it.

PRAs are a special class of disclosure attacks that target CIDSs which publish their alert data publicly over the Internet. Even though the majority of such systems, in this category, exhibit a centralized architecture, e.g., [19, 21], the applicability of the PRAs is agnostic to the architecture of the CIDS [17]. The only requirement is the ability to access the alerts generated by the CIDS. This is usually achieved by either a web front-end (e.g., similar to Figures 1 and 2) or via the utilization of an Application Programming Interface (API).

PRAs were introduced by Lincoln et al. [11] and were further analyzed by several researchers, e.g., [3, 16, 17, 2]. Below, a summary of the idea of the PRA is given along with a brief description of the improvement mechanisms that we proposed in [26].

2.1. Basic PRA logic

In the following, the basic concept of a PRA is introduced, following the attack logic as proposed by Bethencourt et al. [3]. An overview of the lifecycle of such an attack is also given in Figure 3

DOI: 10.1002/nem

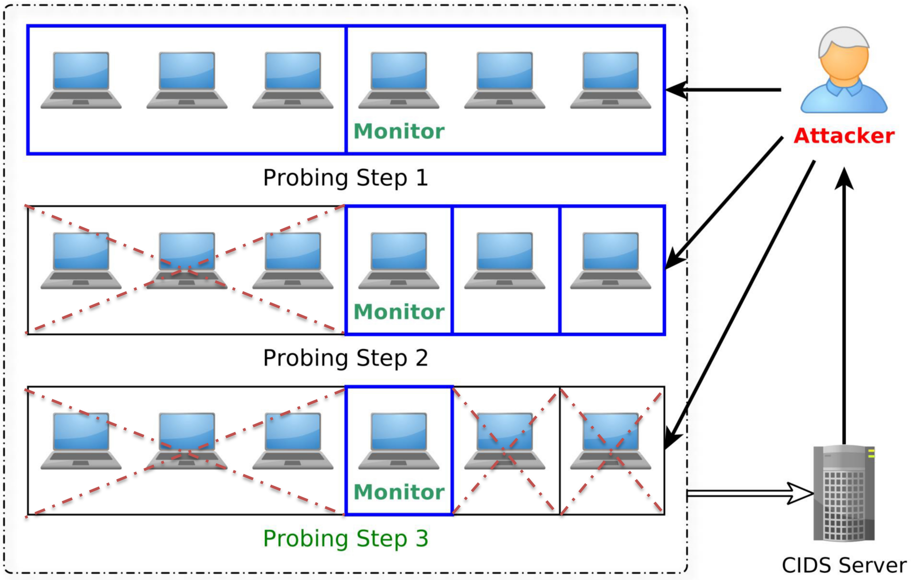
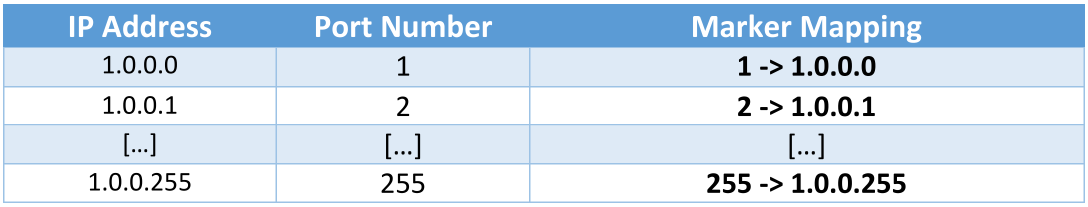


Figure 3. Probe-Response Attack (PRA) lifecycle overview

[25]. The attack involves several steps. First, the adversary begins the PRA by dividing the whole IPv4 address space into equally sized groups (for the sake of simplicity, Figure 3, initially assumes a total of six hosts divided into two groups). Each group is assigned a distinct specially crafted watermark, also known as *marker*[†]. A marker can take many forms; for instance, the adversary can use an uncommon source port to afterwards distinguish the marker from the responses received from the CIDS. A simplistic example[‡] of mapping ports to IP addresses can be seen in Table I.



IP Address	Port Number	Marker Mapping
1.0.0.0	1	**1 -> 1.0.0.0**
1.0.0.1	2	**2 -> 1.0.0.1**
[...]	[...]	[...]
1.0.0.255	255	**255 -> 1.0.0.255**

Table I. Example of marker mapping

[†] This implies that every host inside a group will be tagged with the same marker.

[‡] For the sake of simplicity the example makes use of low port numbers that are usually either reserved or uncommon. The attacker can, of course, make use of more realistic values.

Subsequently, the attacker will probe each host with the respective marker. If a monitor is present among the probed hosts, it will classify the probe as an attack and notify the corresponding CIDS server. The CIDS will publish this incident in its publicly available report. By inspecting the CIDS's published reports, the attacker can determine to which group the monitor belongs to, by examining the respective marker. Afterwards, the adversary carries on with the attack by sending probes to the respective address space.

At a glance, the driving idea behind such a *divide and conquer* attack is that the markers can be then utilized for examining the output of the CIDS and determining whether it contains signs of the markers or not. In this context, and with respect to the received output from the CIDS, the attacker can reduce the probed IP space and repeat the probing steps until the monitors' addresses are revealed.

Bethencourt et al. presented a PRA that follows the aforementioned logic, along with algorithms for efficient probing [3]. In addition, the authors described a variety of adversarial models with regard to the capabilities of the attacker, e.g., the available bandwidth. Bethencourt et al. provided results of various simulations that demonstrate that their PRAs are feasible within a relatively short time-frame. The trade-off, however, in their approach was the bandwidth. On the one hand, with a network speed of 384Mbits/s, 3 days are required to conduct a complete PRA. On the other hand, with a network speed of 1.544Mbits/s, 34 days are required.

2.2. *Improved PRA*

For PRAs to be practically realized, there is a need for efficient and rapid Internet-wide probing. The assumption behind such attacks is that a CIDS utilizes a large number of reachable monitors that are distributed all over the IPv4 address space. Over the last years, research in this domain has made important improvements, e.g., [7, 12]. In particular, Durumeric et al. [7] presented ZMap, a tool for performing Internet-wide network scanning. ZMap significantly reduces the required time for an Internet-wide probing, under certain assumptions, to one hour or even less.

DOI: 10.1002/nem

In our previous work [25, 26], we have presented significant improvements to the speed of the attacks by utilizing such state-of-the-art techniques in Internet-wide probing and by improving the PRAs themselves. In particular, we proposed the Generic Marker Encoding Methodology (GMEM) that combines all available *marker* values and introduces the concept of *checksums*. Checksums solve the problem of noise, i.e., attacks that appear in a CIDS and are mistakenly interpreted (by the adversary) as part of the PRA. This is achieved by introducing a checksum field inside the marker[§]; eventually, when the attacker examines the CIDS output, to be considered part of the PRA, all markers need to comply with the pre-computed checksum. This mechanism can be implemented with various ways, including checksum algorithms or even symmetric encryption mechanisms. For instance, in [26] we utilized the Fletcher checksum algorithm [8].

The proposed methodology offers two major advantages. First, via the utilization of checksums the adversary can reduce the number of repetitions of the PRA and thus reduce the overall execution time. Second, the checksum approach efficiently deals with noise, a problem that had not been efficiently tackled in related work. Our results, in real-world CIDSs, showed that PRAs can be practically executed in less than a day [26]. As the aforesaid mechanisms, i.e., GMEM, emphasize on the improvements of PRAs (from the attackers' perspective) they are considered out of the scope of this article. For a detailed description of GMEM and the checksum idea the reader can refer to [26].

3. DETECTION OF ATTACKS

The first step to cope with PRAs is to detect their presence in a CIDS. In [26], we proposed a statistical anomaly detection technique that is based on the following assumptions. First, in a generic CIDS scenario the adversary has no knowledge of either the IP addresses of the monitors

[§] The checksum is, as the name implies, an encoded value of the marker that is placed inside the PRA message (see also [26]).

nor the exact amount of them. In practice, this is realized by the need, of the adversary, for a large-scale probing, e.g., of the whole IPv4 address space. Second, as a consequence of the aforesaid assumption, it can also be expected that a large amount of monitors will be triggered during a PRA.

In other words, the attacker will be attempting to query the IP address space for signs of CIDS monitors as fast as possible (nowadays this can be achieved within a few hours); during this rather short time frame almost all of the CIDS monitors will trigger alerts. Therefore, the following statistical properties and assumptions are expected during PRAs:

- AS_1: In a certain time-window, the set of reporting monitors that is generating alerts is significantly increased.
- AS_2: The number of unique destination (and/or source) ports will increase (assuming probes are sent out using port-based markers).
- AS_3: The total number of generated alerts in the CIDS increases, but not significantly (i.e., the CIDS generates a large number of alerts regardless of the presence of a PRA).

A real-world connection and reasoning with regard to the aforesaid assumptions is given in the following. First, AS_1 is based on the fact that *i*) a PRA can be nowadays conducted rapidly (see also Section 2.2) and *ii)* not all monitors generate alerts in a very small window of time. Second, assuming that PRAs are likely to utilize ports as markers (cf. [25, 26]), AS_2 holds, as the attacker will be utilizing destination (and/or source) ports as the main watermark for the PRA. Lastly, the reasoning behind AS_3 is straightforward. CIDSs monitor very large networks and hence constantly produce a large amount of alert data. In this context, the additional alerts, that are generated due to the PRAs, are, statistically speaking, not significant enough compared to the total number of alerts.

3.1. Alert-increase based detection

Bearing AS_1 and AS_3 in mind, we proposed an effective metric to detect such attacks by utilizing the *ratio* of generated alerts in relationship to the number of actively reporting monitors. Let A be the set of all generated alerts, S be the set of all monitors, $S_t \subset S$ the set of reporting monitors within time-frame t, and $A_t \subset A$ the set of generated alerts within time-frame t. The ratio r_a is defined as:

$$r_a = \frac{|A_t|}{|S_t|}. \tag{1}$$

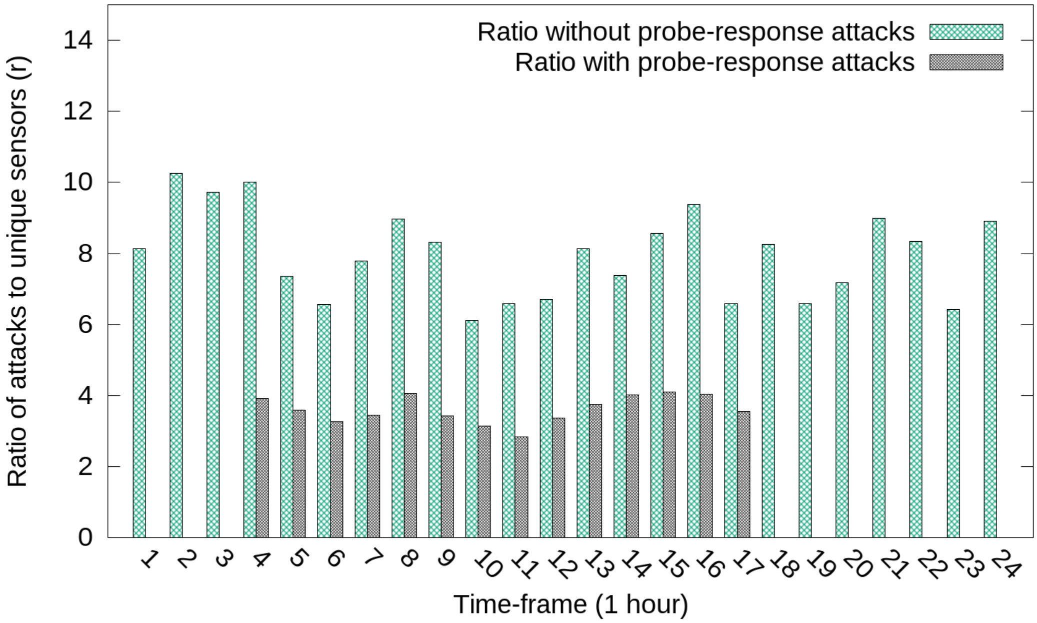


Figure 4. Ratio r_a utilization example for DShield data

To better clarify the meaning and the applicability of this metric, we presented an emulated PRA scenario in which the respective values are calculated with data gathered from the DShield CIDS [26]. Figure 4, depicts how the r_a parameter behaves within a period of 24 hours for data gathered by DShield. In this emulated PRA scenario, it is assumed that an attacker requires approximately 5 hours (assuming a 100Mbit/s network connection) to perform one probing step in the entire IPv4 range [7], probing approximately $90,000$ monitor addresses per hour (assuming a total of $500,000$ monitors[¶]). With respect to the aforementioned assumptions, in the presence of a PRA the number of (unique) reporting monitors within a time-frame $|S_t|$ will increase significantly (cf. assumption AS_1), while $|A_t|$ will only have a relatively small increase (cf. assumption AS_3),

[¶] This is an approximation of the total number of monitors that the DShield CIDS possesses.

therefore modifying r_a. In the presented period one can observe the monitors $|S| = 131,344$, the alerts $|A| = 10,934,768$, and an average unique monitor count (per hour) $\sum_t \frac{|S_t|}{24} = 55,000$.

A PRA was emulated by introducing alarms in the time-frames between 4 and 17 in a 24 hour period. By assuming that the maximum probing rate is $90,000$ and that monitors might already be present, the PRAs are injected according to a uniform distribution between $80,000$ and $90,000$. As it is depicted in Figure 4, it becomes evident that during an attack the ratio r_a decreases significantly. The reason for this decrease is connected to assumptions AS_1 and AS_3. That is, while the number of unique monitors (generating alerts) in a specific time frame significantly increases, the increase in the total number of alerts is only minor.

3.2. *Port-based detection*

Another technique for detecting the presence of PRAs is by studying the *frequency* of unique destination ports in a specific time-window [26]. In contrast to source ports (that are usually chosen randomly), destination ports can be utilized as markers and thus their number is expected to increase during a PRA. In this case it is important to carefully decide which time window should be taken for studying the respective port frequency.

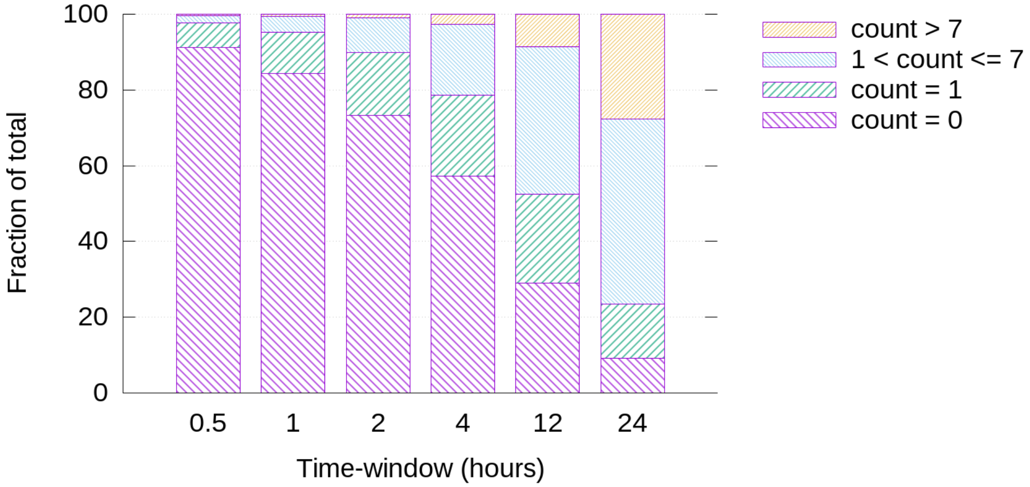


Figure 5. Destination port frequency for different time-windows in DShield

DOI: 10.1002/nem

Figure 5, shows the distribution of port frequency in DShield by setting a fixed start time and extending the window up to 24 hours. As one can observe, in the first half hour almost 93% of the ports are not utilized, while when the window is increased this percentage is decreased rapidly. This suggests that large time-windows (e.g., more than two hours) might introduce many false positives.

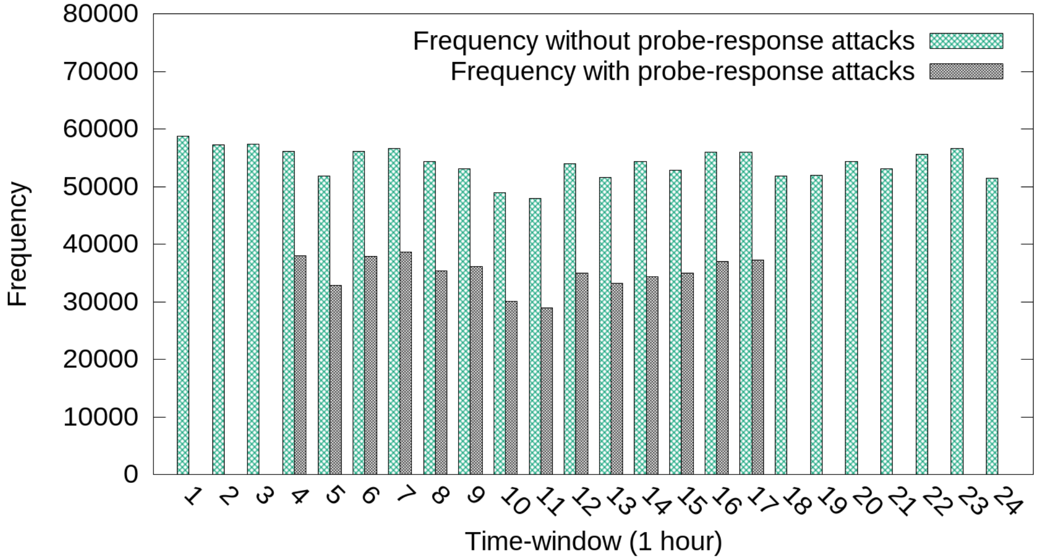


Figure 6. Destination port frequency in DShield

Bearing this in mind, Figure 6, with a similar setup as Figure 4, shows how the frequency (of destination ports in a specific time-window) evolves under the presence and absence of an emulated PRA. Hence, the difference between attacked and non-attacked states can be utilized as a threshold for the detection of PRAs.

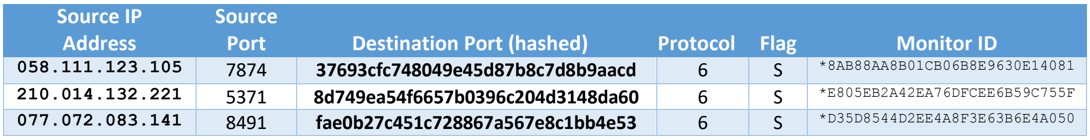
4. MITIGATION OF ATTACKS

In this section, all the prominent proposed mitigation techniques are discussed. Note that the mitigation mechanisms that are described in this section assume a method for the detection of a PRA (cf. the previous section).

DOI: 10.1002/nem

4.1. *Hashing*

As the name implies *hashing*, in the context of PRA mitigation, refers to the process of utilizing a hash function to map marker values. By doing so, the respective parameters become unusable from the adversaries' perspective, thus reducing the applicability of the PRA. This defense mechanism was first proposed by Bethencourt et al. [3]. An example of the utilization of the hashing mechanism is shown in the third column of Table II. Specifically, in this example the destination port values have been hashed via the utilization of the MD5 hash function. Hence, when an adversary attempts to use the destination port as a marker while performing a PRA, he will not succeed as all the respective values are hashed.



Source IP Address	Source Port	Destination Port (hashed)	Protocol	Flag	Monitor ID
058.111.123.105	7874	37693cfc748049e45d87b8c7d8b9aacd	6	S	*8AB88AA8B01CB06B8E9630E14081
210.014.132.221	5371	8d749ea54f6657b0396c204d3148da60	6	S	*E805EB2A42EA76DFCEE6B59C755F
077.072.083.141	8491	fae0b27c451c728867a567e8c1bb4e53	6	S	*D35D8544D2EE4A8F3E63B6E4A050

Table II. Example of hashing the destination port marker in the DShield data

There are a number of shortcomings with regard to hashing. First, hashing seriously damages the usability of the dataset making it unreadable for the legitimate users. Furthermore, in the case in which the mitigation mechanism is activated only upon detection of a PRA, the adversary will immediately realize that the attack has been identified. Similarly, a malicious entity may choose to utilize such knowledge to enforce the CIDS to perform hashing and thus reduce its usability. Lastly, when a known hash function (instead of an encryption scheme such as symmetric encryption - cf. Section 4.2, below) is utilized for a specific range of integer values (e.g., port numbers) an attacker may use a rainbow-table like technique to create a database of all possible values. Correspondingly the adversary can reverse the hash values.

4.2. *Encryption*

Encryption refers to a similar approach to hashing; in this case, instead of utilizing a hash function the defender can instead encrypt the marker values by either using symmetric or

DOI: 10.1002/nem

asymmetric cryptographic techniques. Encrypting the markers can assist for overcoming some of the aforementioned disadvantages of hashing. However, it does not offer a solution for the usability trade-off. In addition, other challenges arise; for instance, what type of cryptographic techniques should be used, how would the keys be distributed in the CIDS, etc.

4.3. Adaptive Sampling

Sampling was first mentioned in [3] and was further improved and analyzed by us in [26, 25]. In more details, the idea behind this mitigation method is that the CIDS will selectively publish only a sample of the overall generated attacks whenever it detects the presence of a PRA. The intensity of the sampling can also be proportional to the attack intensity [26]. Therefore, the attacker will not be able to retrieve all the marker probes from the CIDS. This leads to a reduction of the effectiveness of the attack. Our simulation results suggested that, by utilizing this adaptive sampling approach, only the 31% of the total monitors were detected by the PRA [26]. However, as a result of the sampling process, there is also a reduction of 62% in the total number of events that are reported by the CIDS.

The main shortcoming of sampling is the impact that it has in the usability of the CIDS. That is, the system publishes only a small portion of the overall alert data. In fact, an adversary might attempt to exploit this mechanism and perform a type of a Denial of Service (DoS) attack on the system. Moreover, the trade-off between effectiveness and usability is not very satisfying as a 31% PRA success rate is rather high.

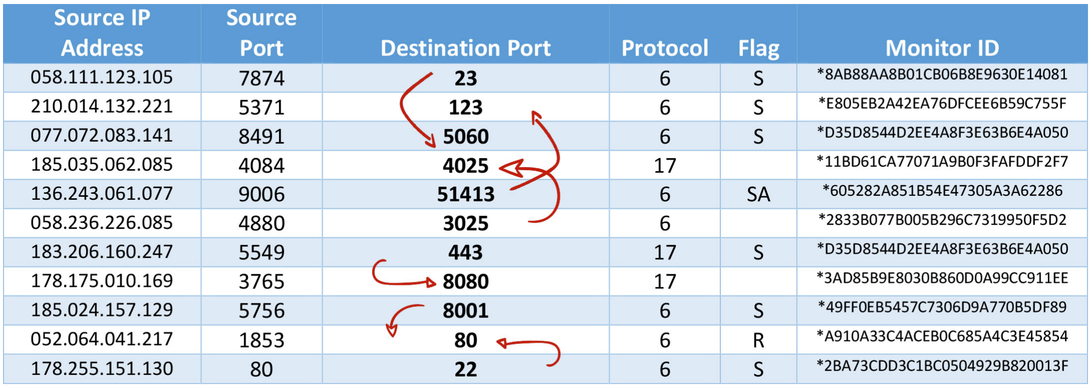
4.4. Shuffling-based PRA Mitigation

The idea behind the Shuffling-based PRA mitigation is based on the fact that only a few of the parameters in the publicly available output of CIDS can be practically utilized as probe markers [25]. These possible markers can be easily anticipated by carefully examining the CIDS. For instance, Figure 2 depicts some of the output parameters of the TraCINg CIDS from which one can derive all possible probe markers (e.g., the destination port).

Based on this observation, in [24], we proposed the Shuffle-based PRA Mitigation (SPM). In SPM the defender shuffles, i.e., changes the positions, of certain parameters upon detection of a

DOI: 10.1002/nem

PRA. This concept is inspired by the *shell game*, a deception approach that has also been used as a state of the art technique for achieving anonymity [5]. With this we attempt to bridge the trade-off between effectively defending against PRAs and, by doing so, reducing the usability of the CIDS.



Source IP Address	Source Port	Destination Port	Protocol	Flag	Monitor ID
058.111.123.105	7874	**23**	6	S	*8AB88AA8B01CB06B8E9630E14081
210.014.132.221	5371	**123**	6	S	*E805EB2A42EA76DFCEE6B59C755F
077.072.083.141	8491	**5060**	6	S	*D35D8544D2EE4A8F3E63B6E4A050
185.035.062.085	4084	**4025**	17		*11BD61CA77071A9B0F3FAFDDF2F7
136.243.061.077	9006	**51413**	6	SA	*605282A851B54E47305A3A62286
058.236.226.085	4880	**3025**	6		*2833B077B005B296C7319950F5D2
183.206.160.247	5549	**443**	17	S	*D35D8544D2EE4A8F3E63B6E4A050
178.175.010.169	3765	**8080**	17		*3AD85B9E8030B860D0A99CC911EE
185.024.157.129	5756	**8001**	6	S	*49FF0EB5457C7306D9A770B5DF89
052.064.041.217	1853	**80**	6	R	*A910A33C4ACEB0C685A4C3E45854
178.255.151.130	80	**22**	6	S	*2BA73CDD3C1BC0504929B820013F

Table III. Example of the shuffling procedure in DShield data

The shuffling in SPM is stochastic and can be realized via the utilization of a pseudo-random function. As expected, due to the stochastic nature of the process and the limited range of the parameters (e.g., ports can have 65537 possible values) there will be cases in which the result of the shuffling process will be the same with the original parameter. However, as it will be shown below (see Section 4.6 and Figures 7 and 8), only a very small portion of the monitors can be detected as result of this. Table III, illustrates an example of the SPM procedure in the case of DShield data when adjusting the destination port value. Note that for the sake of visual clarity the table depicts the shuffling process only for some of the parameters.

The main advantage of the SPM approach is that it requires minimal modifications in the alert data. Note that as the data is shuffled, but not altered, global statistics will still be valid (e.g., creating lists of most commonly attacked ports or protocols). Moreover, the adversary cannot know if the CIDS has detected the presence of the PRA and/or whether the system has activated defense measures. This is not the case with other methods such as the hashing and the encryption of the alert data.

DOI: 10.1002/nem

4.5. Other approaches

There have been proposed some additional approaches for the mitigation of PRAs that, however, require either a dramatic reduction of the usability of the system or an overwhelming overhead for the administrators.

4.5.1. Non-Public CIDS: A simplistic approach for completely canceling the ability to perform a PRA is by cutting out the feedback loop. This can be easily done by making the CIDS private, e.g., by enforcing access control into its contents [3]. Nevertheless, such an approach completely disregards the benefits of sharing alert data publicly.

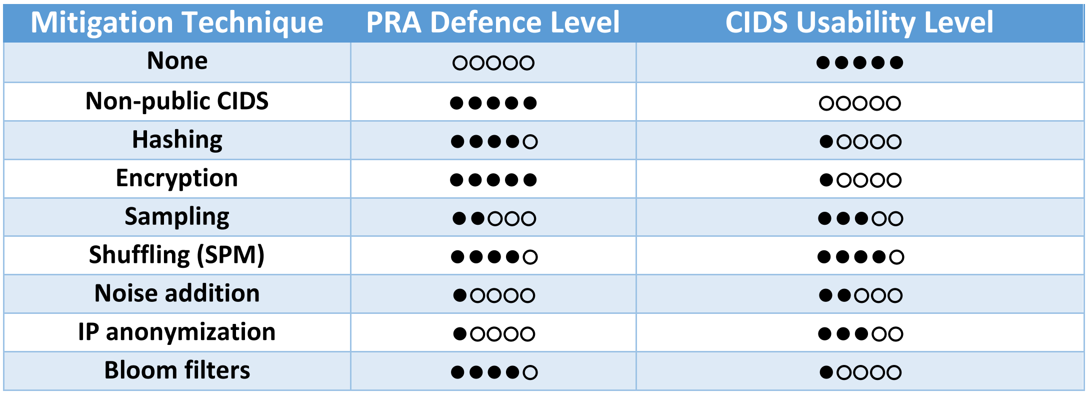
4.5.2. Dynamic IP addresses: Another approach is to regularly change the network position, i.e., the IP address, of the monitors [3]. Such an approach would effectively tackle the PRA problem but it would also introduce massive overhead for the administrators of the system. For instance, in the DShield CIDS there are approximately $500,000$ monitors, which, in their majority, are managed by organizations outside DShield. Furthermore, in many organizations the range of IP address (especially with regard to IPv4) is very limited.

4.5.3. Prefix IP anonymization: Bethencourt et al. proposed the utilization of anonymization techniques [27] for the IP addresses in CIDSs that publish such data (and hence can be utilized as PRA markers) [3]. However, this does not solve the PRA problem as the adversary can utilize different types of alert data as markers (e.g., ports).

4.5.4. Bloom filter utilization: Bethencourt et al., also discussed the usage of bloom filters in the sense of reducing the available marker surface for the attackers [3]. In general, bloom filters can indeed be of benefit for CIDSs [10, 23] but for different purposes (e.g., anonymization of data or reducing the communication overhead). Nevertheless, this creates the same problems with encrypting and hashing since the usefulness of the CIDS is highly affected and also because bloom filters cannot be practically published as meaningful output from the CIDS.

DOI: 10.1002/nem

4.5.5. Noise addition: Finally, another method is the addition of *noise* data in the CIDS output. Based on the specifics of the noise data this approach can mitigate the original PRA [3]. However, such an approach cannot defend against the more sophisticated PRAs [26] due to the watermark that is included inside the markers. In addition, this method contaminates the alert data.



Mitigation Technique	PRA Defence Level	CIDS Usability Level
None	○○○○○	●●●●●
Non-public CIDS	●●●●●	○○○○○
Hashing	●●●●○	●○○○○
Encryption	●●●●●	●○○○○
Sampling	●●○○○	●●●○○
Shuffling (SPM)	●●●●○	●●●●○
Noise addition	●○○○○	●●○○○
IP anonymization	●○○○○	●●●○○
Bloom filters	●●●●○	●○○○○

Table IV. Comparison of different PRA mitigation techniques: ”○ ○ ○ ○ ○” indicates the lowest (worst-case) possible value, while ”● ● ● ● ●” the highest (best-case) one

4.6. Summary

We argue that there is a trade-off between defending against a PRA and maintaining the usability of the CIDS. Therefore, the respective research challenge is to identify mechanisms that, on the one hand, disrupt the PRA process while, on the other hand, introduce minimal or zero overhead on the operation of the system.

Table IV, summarizes the analysis of this section. In particular, it compares all the different mitigation mechanisms presented in Section 4 with regard to the PRA defense level and the overall CIDS usability level (after the implementation of the respective measure). The comparison here is *qualitative* and follows the argumentation of the article. Nevertheless, the findings of Table IV, and specifically with regard to the PRA defense level, correspond to the simulation results as shown in Figures 7 and 8 [24].

In more details, Figures 7 and 8 depict a comparison of *sampling*, *hashing* and *shuffling* (SPM) (see also Section 4.4). Particularly, the figures show how many monitors have been identified in

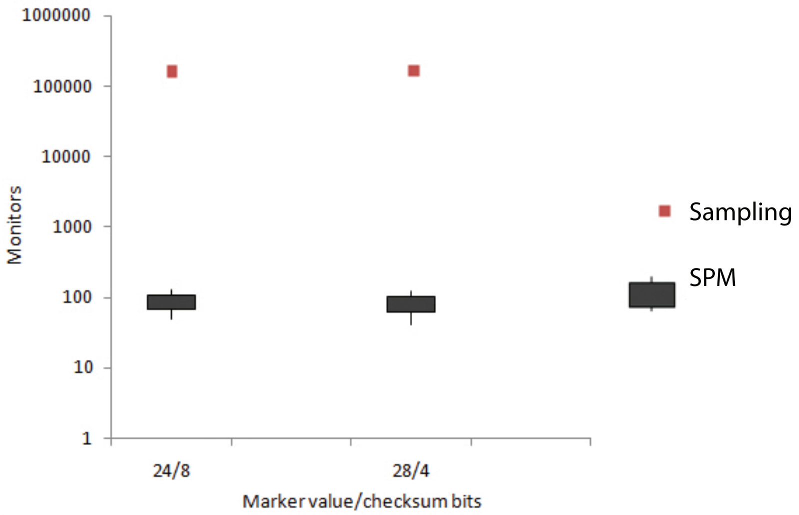


Figure 7. Comparison of the shuffling (SPM) mitigation technique and sampling

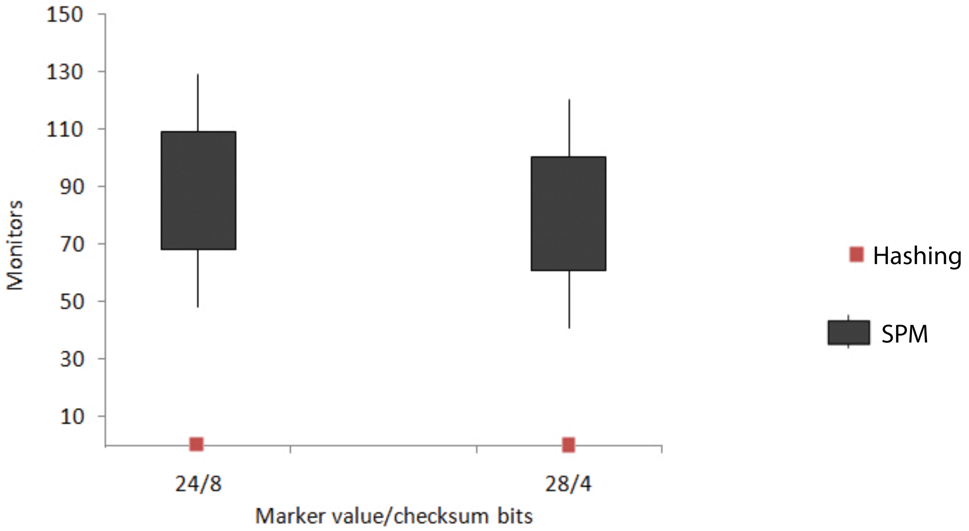


Figure 8. Comparison of the shuffling (SPM) mitigation technique and hashing

a CIDS with $500,000$ monitors, which is attacked by a PRA in a simulation setup. As expected, hashing performs best (with a trade-off of usability – see also Section 4.1) while shuffling performs also good (without the aforementioned strong usability trade-off). Lastly, sampling performs poorly compared to the other two methods. We refer the reader to [24], for a detailed description of the simulation setup of the aforesaid experiments. Finally, measuring the *usability* level of a CIDS,

DOI: 10.1002/nem

while deploying PRA defense mechanisms, in an unbiased manner is a challenging task and is considered out of the scope of this article.

5. FUTURE WORK AND RESEARCH GAPS

In this section, we discuss research gaps, that we identified in the areas of detecting and mitigating PRAs, and propose ideas for future work.

5.1. Detecting PRAs

The detection of PRAs is not a very mature or well-studied area. In fact, besides the methods shown in Section 3 the majority of mitigation techniques assumes the ability to detect PRAs. Bearing this in mind, in the following, we attempt to identify gaps and future directions in this area.

- Some assumptions regarding the detection metrics (see Section 3) might not always hold. For instance, a highly targeted attack might not target the whole IPv4 address space. This implies that if an adversary is attempting to, for example, detect monitors in a particular country (by only scanning the respective IP space) then the assumption regarding the increase in the number of monitors will not hold (see assumption AS_1 in Section 3).
- A similar problem may occur with the assumption AS_2 (Section 3) when the adversary is not making use of ports as the main marker.
- Defining a meaningful threshold value is critical for the ratio-based detection methods, presented in Section 3. In our previous work, we manually derived the threshold values by analyzing historical data. However, this requires a lot of manual work and is prone to generating false negatives and false positives. To overcome the aforesaid limitations, machine learning techniques can be utilized for deriving proper threshold values [1]. Such a method will have the additional advantage of being agnostic of the CIDS.
- A hybrid technique that combines the two detection methods (presented in Section 3) is likely to decrease/minimize false positives. In more details, the defender can utilize the ratio r_a

DOI: 10.1002/nem

(see Equation 1) in combination to measurements with regard to the frequency of unique port destinations to have a more holistic detection mechanism.

5.2. *Mitigating PRAs*

In contrast to the detection of PRAs, several methods have been proposed for their mitigation (see Section 4). We envision the following additional ideas and improvements towards a more effective mitigation of such attacks.

- For a plethora of reasons the utilization of $IPv6$ can significantly help for the mitigation of PRAs. For instance, the total number of possible $IPv6$ addresses is considerably larger than $IPv4$ (from 2^{32} to 2^{128}), which would increase the time required to perform a full scan of the address space to a non-feasible level.
- The main assumption behind a successful PRA (see Section 2) is that the monitor of the CIDS will classify the probe as an attack and notify the corresponding CIDS server. Afterwards, the CIDS will publish the respective alert, providing a feedback loop for the adversary. Therefore, a novel approach for mitigating such attacks is to attempt to create signatures for the probe messages. This might not be trivial but it is definitely feasible. For instance, Doerr et al, showed techniques for the detection of of ZMap scans based on the IP selection behavior of ZMap [6]. Since ZMap is nowadays the basis for the execution of PRAs [26], this can be exploited for the defenders' benefit.
- With regard to the Shuffle-based PRA Mitigation (SPM) technique, it might be possible to utilize pseudo-random generators of which the utilized seeds can be shared with trusted users so that the whole process can be reversible. Note that, this would not influence the effectiveness or the security of the mechanism (against the PRA) since the adversary cannot predict whether the SPM is activated in a certain time-window or not.
- As mentioned is Section 4.6 the *CIDS usability level* is an important parameter that is, however, not easy to quantify in an unbiased manner. Thus, further work is required towards such a task; for instance, by performing user studies.

DOI: 10.1002/nem

6. CONCLUSION

Probe-Response Attacks (PRAs) introduce a threat to a Collaborative Intrusion Detection System (CIDS) by allowing malicious entities to detect the network location (IP addresses) of the monitors of the system. In this article we discussed the state of the art with respect to the proposed detection and mitigation techniques. As we have seen in the previous sections, detecting PRAs is possible when certain assumptions hold. In particular, even though the existing methods for detecting the presence of a PRA seem to be effective, we argue that more work is required for a completely CIDS-agnostic and automated method. Defending against and mitigating such attacks is not an easy task and by examining the state of the art a trade-off was identified between successful mitigation and the usability of the CIDS's output after the implementation of the defensive measures. This can be the basis for further improvements in PRA mitigation. Lastly, in this article we presented various research gaps for both the detection and mitigation of PRAs.

REFERENCES

1. Mohiuddin Ahmed, Abdun Naser Mahmood, and Jiankun Hu. A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, 60:19–31, 2016.
2. Paul Barford, Somesh Jha, and Vinod Yegneswaran. Fusion and filtering in distributed intrusion detection systems. In *Proc. Allerton Conference on Communication, Control and Computing*, 2004.
3. John Bethencourt, Jason Franklin, and Mary Vernon. Mapping internet sensors with probe response attacks. In *USENIX Security Symposium*, pages 193–208, 2005.
4. Leon Böck, Emmanouil Vasilomanolakis, Max Mühlhäuser, and Shankar Karuppayah. Next generation p2p botnets: Monitoring under adverse conditions. In *Research in Attacks, Intrusions, and Defenses (RAID)*, pages 511–531. Springer International Publishing, 2018.
5. Jörg Daubert, Mathias Fischer, Tim Grube, Stefan Schiffner, Panayotis Kikiras, and Max Mühlhäuser. Anonpubsub: Anonymous publish-subscribe overlays. *Computer Communications*, 76:42–53, 2016.
6. C. Doerr, M. el Maouchi, S. Kamoen, and J. Moree. Scan prediction and reconnaissance mitigation through commodity graphics cards. In *2016 IEEE Conference on Communications and Network Security (CNS)*, pages 287–295, Oct 2016.
7. Zakir Durumeric, Eric Wustrow, and J. Alex Halderman. ZMap: Fast Internet-wide Scanning and Its Security Applications. In *Proceedings of the 22nd USENIX Security Symposium*, pages 605–619, 2013.

DOI: 10.1002/nem

8. John G Fletcher. Arithmetic checksum for serial transmissions. *IEEE Transactions on Communications*, (1):247–252, 1982.
9. Carol Fung. Collaborative intrusion detection networks and insider attacks. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 2(1):63–74, 2011.
10. Philip Gross, Janak Parekh, and Gail Kaiser. Secure selecticast for collaborative intrusion detection systems. In *Proceedings of the 3rd International Workshop on Distributed Event-Based Systems (DEBS'04)*. IET, 2004.
11. Patrick Lincoln, Phillip A. Porras, and Vitaly Shmatikov. Privacy-preserving sharing and correction of security alerts. In *13th USENIX Security Symposium*, pages 239–254, 2004.
12. Dirk Maan, José Jair Santanna, Anna Sperotto, and Pieter-tjerk De Boer. Towards validation of the Internet Census 2012. In *20th EUNICE/IFIP EG 6.2, 6.6 International Workshop*, pages 85–96. Springer, 2014.
13. Robert Mitchell and Ing-Ray Chen. A survey of intrusion detection techniques for cyber-physical systems. *ACM Computing Surveys (CSUR)*, 46(4):55, 2014.
14. David Moore, Colleen Shannon, and Jeffery Brown. Code-Red: A Case Study on the Spread and Victims of an Internet Worm. In *Second ACM SIGCOMM Workshop on Internet Measurment (IMW)*, pages 273–284, 2002.
15. David Moore, Colleen Shannon, Geoffrey M Voelker, and Stefan Savage. *Network telescopes: Technical report*. Department of Computer Science and Engineering, University of California, San Diego, 2004.
16. Yoichi Shinoda, Ko Ikai, and Motomu Itoh. Vulnerabilities of passive internet threat monitors. In *USENIX Security Symposium*, pages 209–224, 2005.
17. Vitaly Shmatikov and Ming-Hsiu Wang. Security against probe-response attacks in collaborative intrusion detection. In *Workshop on Large scale attack defense - LSAD*, pages 129–136, New York, USA, 2007. ACM.
18. Aditya K. Sood and Richard J. Enbody. Targeted Cyber Attacks-A Superset of Advanced Persistent Threats. *IEEE Security & Privacy*, 11(1):54–61, 2013.
19. Johannes Ullrich. Dshield internet storm center. https://www.dshield.org/, 2000.
20. Emmanouil Vasilomanolakis. *On Collaborative Intrusion Detection*. PhD thesis, Technische Universität Darmstadt, Darmstadt, July 2016.
21. Emmanouil Vasilomanolakis, Shankar Karuppayah, Panayotis Kikiras, and Max Mühlhäuser. A honeypot-driven cyber incident monitor: lessons learned and steps ahead. In *International Conference on Security of Information and Networks*, pages 158–164. ACM, 2015.
22. Emmanouil Vasilomanolakis, Shankar Karuppayah, Max Mühlhäuser, and Mathias Fischer. Taxonomy and Survey of Collaborative Intrusion Detection. *ACM Computing Surveys*, 47(4):33, 2015.
23. Emmanouil Vasilomanolakis, Matthias Krügl, Carlos Garcia Cordero, Max Mühlhäuser, and Mathias Fischer. Skipmon: A locality-aware collaborative intrusion detection system. In *Computing and Communications Conference (IPCCC), 2015 IEEE 34th International Performance*, pages 1–8. IEEE, 2015.
24. Emmanouil Vasilomanolakis, Noorulla Sharief, and Max Mühlhäuser. Defending against probe-response attacks. In *IEEE/IFIP Workshop on Security for Emerging Distributed Network Technologies (DISSECT)*. IEEE, 2017.

DOI: 10.1002/nem

25. Emmanouil Vasilomanolakis, Michael Stahn, Carlos Garcia Cordero, and Muhlhauser Max. Probe-response attacks on collaborative intrusion detection systems: Effectiveness and countermeasures. In *Communications and Network Security (CNS)*, pages 699–700. IEEE, 2015.

26. Emmanouil Vasilomanolakis, Michael Stahn, Carlos Garcia Cordero, and Max Mühlhäuser. On probe-response attacks in collaborative intrusion detection systems. In *Conference on Communications and Network Security (CNS). IEEE*, 2016.

27. Jun Xu, Jinliang Fan, Mostafa H Ammar, and Sue B Moon. Prefix-preserving ip address anonymization: Measurement-based security evaluation and a new cryptography-based scheme. In *Network Protocols, 2002. Proceedings. 10th IEEE International Conference on*, pages 280–289. IEEE, 2002.

28. Chenfeng Vincent Zhou, Christopher Leckie, and Shanika Karunasekera. A Survey of Coordinated Attacks and Collaborative Intrusion Detection. *Computers & Security*, 29(1):124–140, feb 2010.

DOI: 10.1002/nem