Fundamental Scaling Laws of Covert DDoS Attacks

Amir Reza Ramtin, Philippe Nain, Daniel Sadoc Menasche, Don Towsley, Edmundo de Souza

Fundamental Scaling Laws of Covert DDoS Attacks

Amir Reza Ramtin[a,*], Philippe Nain[b], Daniel Sadoc Menasche[c], Don Towsley[a], Edmundo de Souza e Silva[c]

[a] *University of Massachusetts, Amherst, USA {aramtin,towsley}@cs.umass.edu*
[b] *Inria, Sophia Antipolis, France, philippe.nain@inria.fr*
[c] *Federal University of Rio de Janeiro, Brazil, sadoc@dcc.ufrj.br, edmundo@land.ufrj.br*

Abstract

Botnets such as Mirai use insecure home devices to conduct distributed denial of service attacks on the Internet infrastructure. Although some of those attacks involve large amounts of traffic, they are generated from a large number of homes, which hampers their early detection. In this paper, our goal is to answer the following question: what is the maximum amount of damage that a DDoS attacker can produce at the network edge without being detected? To that aim, we consider a statistical hypothesis testing approach for attack detection at the network edge. The proposed system assesses the goodness of fit of traffic models based on the ratio of their likelihoods. Under such a model, we show that the amount of traffic that can be generated by a covert attacker scales according to the square root of the number of compromised homes. We evaluate and validate the theoretical results using real data collected from thousands of home-routers connected to a mid-sized ISP.

Keywords: Hypothesis testing, Scaling laws, Gaussian mixture, Covertness, DDoS attack, Home networks

1. Introduction

The Internet has become an indispensable commodity in the last several years. This achievement was parallel to the growth of sophistication that home

*Corresponding author

networks have undergone, nowadays hosting a variety of devices such as PCs, tablets, mobile phones and specialized apparatus such as smart thermostats and other Internet of things (IoT) devices. While these devices offer users an array of services and conveniences, they come at the cost of increasing the attack surface of the home network [1, 2, 3]. Because of the vulnerabilities of such devices, they have been increasingly used as the source of Distributed Denial-of-Service (DDoS) attacks [4]. According to the *European Union Agency for Cybersecurity* the number of DDoS attacks has increased significantly in 2020 and the trend continues [5]. These attacks are most harmful to services, and very costly to organizations, both in terms of time and money, since they may cripple key system's resources.

DDoS attacks are difficult to prevent, because they are launched from a large number of infected devices connected to the Internet, collectively known as botnets. The attacker compromises devices by injecting malicious code (malware), which allows the attacker to perform actions at a later time using these devices as sources of harmful traffic without knowledge of the device's owner. The traffic generated by some botnets is typically composed of millions of small flows [1].

Despite all the continuing efforts to detect and mitigate these attacks, their number have not decreased and it has been predicted that this number will double from 2018 to 2023 [6]. In fact, the number of DDoS attacks drastically increased in 2020 [7]. Roughly, DDoS attacks are produced by launching a burst of packets simultaneously from a very large number of devices towards a given target. Examples of common attack types include: (a) UDP-flood attacks: ports of a remote host are flooded with UDP packets which can cripple the target by draining resources to process the arriving packets; (b) ICMP-flood attacks: the target is flooded with ICMP packets as fast as possible to produce a response from the target which, in turn, may cause a considerable system slowdown; (c) SYN-flood attacks: the victim is flooded with TCP SYN packets and, for each packet received, a SYN-ACK is produced and the target waits for acknowledgement from the source that will never arrive, committing resources for the faked

connection; (d) HTTP-flood attacks, which employ GET or POST requests to a web service. In most cases, the attacker uses either a large number of control packets to overwhelm the victim and exhaust its resources, or packets that do not respect flow control and consume bandwidth resources in the neighborhood of the target.

Needless to say, early identification of these attacks and their sources is of prime concern of companies. However, it is also imperative to discover if there are fundamental tradeoffs between the amount of damage an attacker can inflict to services and the attacker's ability to remain undetected. If these fundamental laws exist, they could shed some light concerning covertness versus damage and they could be used to help building effective DDoS countermeasures.

It should be evident that the objective of the attacker is to inflict as much damage as possible by generating enough traffic (for instance, generating a large amount of control packets) to wear out the victim's resources and, consequently, to disrupt user's services. The malicious traffic originates from home network devices with limited capacity. As such, the attack generated from a single home is far from sufficient to cause any damage. Then, necessarily, the attacker tries to use as many homes as possible, remotely activating a large number of controlled devices (the bots) that have been previously infected. Furthermore, it is advantageous for the attacker to remain *covert* (undiscovered) while attacking.

It is hard to differentiate attack traffic originating from a single home network from the regular home user traffic. This is probably why most network-based DDoS detection methods rely on detailed network traffic information (e.g., packet header data), which is in general computationally expensive and also raises concerns about user privacy. To avoid these drawbacks, methods based solely on metrics such as byte/packet counts should be preferred [8, 9]. A lightweight approach that employs network interface byte/packet counts also scales, and is oblivious to botnet-specific attack signatures and encryption.

Clearly, the larger the number of compromised devices in different home networks, the greater the amount of damage the attacker who controls these bots can potentially cause. The work of [8] proposes a method to detect an

ongoing attack from a home-router without resorting to packet inspection, and also shows that the likelihood of detecting a DDoS attack can be improved with the number of participant bots in the attack. A fundamental question then arises: Can a DDoS attack be covert and if so, what is the damage it is expected to cause?

Goals. We want to avoid packet inspection as in [8]. Furthermore, it should be evident that the damage caused by common DDoS attacks (such as Mirai) is proportional to the number of infected devices (equivalently the number homes) participating in the attack. In addition, from the administrator's point of view, the number of false alarms should be kept to a minimum, since there is no point in detecting occasional attacks if the number of false alarms is unbearably high.

We then pose the following questions related to the attacker's ability to cause as much damage as possible and the likelihood to remain undetected:

1. Is there any fundamental limit on the damage an attacker can cause to the victim while remaining covert?
2. If such limit exists, how is it related to the false alarm rate?

To answer the above questions, we propose an analytical model to capture the essence of these attacks. The model comprises two components, characterizing regular traffic and traffic when an attack is underway. In particular, we focus on the simplest case wherein each component is associated with a single feature, such as byte counts or packet counts observed per time slot.

At a high level, we posit that an attacker is covert if admin running a *detector* (also known as a *classifier*) cannot determine if an attack is in progress by observing the traffic (byte or packet rate) from a set of homes. Formally, consider that admin runs an optimal statistical hypothesis test and uses it to compute the probability of false alarm (p_{FA}) and the probability of miss detection (p_{MD}), both probabilities formally defined in Section 2.2. In this setting, the sum of errors $p_{FA} + p_{MD}$ lies in $[0, 1]$, as shown in Section 2.3. Following the definition in [10], we then say (Definition 2.1) that an attack is covert if the attacker has a strategy that makes the sum $p_{FA} + p_{MD}$ arbitrarily close to one. We stress that

our goal is *not* to devise a deployable detection method but rather to discover fundamental laws that govern the covertness 'game' played between the attacker and admin and to understand the limits on the damage an attacker can cause.

Our model assumptions are backed by real data collected at home-routers from a mid-sized ISP, with whom we partnered to gather statistics about baseline regular traffic. Our dataset includes packets and byte counts collected at thousands of home-routers over several months. Our analysis of the dataset shows that regular traffic can be modeled by a mixture of Gaussian distributions. We also use a dataset of attack traffic, generated by controlled experiments using real botnet code [8]. The traffic distribution of the attack traffic can also be approximately modeled by a mixture of Gaussian distributions.

We establish that the amount of traffic that an attacker can issue while remaining covert grows as $O(\sqrt{n})$, where n is the number of compromised homes controlled by the attacker in the network. We also obtain conditions under which this bound is tight. We confirm these results using the real data mentioned earlier.

Prior art. The covertness criterion considered in this paper was proposed in the context of low probability of detection (LPD) communications. Although there are a number of papers in this area [10, 11, 12, 13, 14], to the best of our knowledge no previous work has discussed hypothesis testing methods for DDoS detection in home networks, particularly focusing on covertness. We are also unaware of prior work analyzing the fundamental laws of covert DDoS attacks through theoretical bounds derived from optimal statistical hypothesis tests.

Contributions. In summary, our contributions are threefold:

- **Analytical model to assess damage:** We develop an analytical model to assess the maximum damage that the attacker can cause from home networks. Our results are based on statistical hypothesis testing under the constraint that the attacker is covert (Section 3);

- **Square root law:** Under the proposed model, we show that the damage caused by the attacker follows a square root law. The amount of mali-

cious packets per time unit a covert attacker can inject during the attack grows as the square root of the number of home-routers in large networks (Section 4);

- **Evaluation:** The main theoretical results are asymptotic when the number of compromised home-routers goes to infinity. Using real traffic traces collected from thousands of home-routers, we show that the asymptotic results are robust when the number of comprimised homes is finite, for different system parameter values. Our evaluation considers two scenarios: in the first, the administrator knows the attack traffic distribution and, in the second, the administrator does not have any knowledge concerning the distribution of the attack traffic (Section 5).

We describe the system under study in Section 2. Sections 3 to 5 follow the outline presented in the summary of contributions above. Section 6 reports related work and Section 7 concludes.

2. System Description and Background

In this section we describe the system that is the focus of our work. We follow this with terminology and basic concepts pertaining to statistical hypothesis tests and covertness.

2.1. System Description

As mentioned in the introductory section, an attacker injects data into the network through previously compromised devices residing in homes. The attacker installs malicious code (malware) at the devices and assumes remote control over them. Examples of such devices include televisions, media stations, etc. Henceforth, we will refer to a home wherein there are compromised devices as a *compromised home*. A set of compromised homes forms a *botnet*.

The attacker controls the botnet, and may use all bots, or a fraction of them, to issue attacks against its target. In the remainder of this work, we focus on

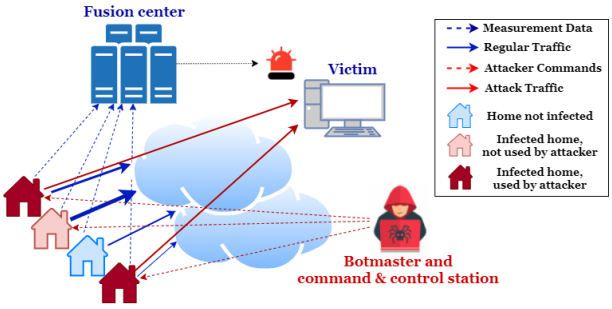


Figure 1: System outline

home networks as the major source of attacks, referring to the compromised homes that can be used by the attacker to issue DDoS attacks simply as *homes*.

The attacker faces the problem of determining which homes to activate and at what rates to inject traffic into the network. *Attack data* is the traffic the attacker transmits through the network, from selected homes to a target. The attack rate is the rate at which the attacker transmits attack data, measured in bytes or packets per second. Usually the attacker pushes as much traffic (UDP packets or control packets) through the infected device's interface as possible to increase the damage a botnet can cause. However, in this work, we also allow the attacker to control the rate at which it injects traffic as an additional option to keep the attack covert. In summary, the attacker has two options – namely, determining the number of homes to activate and/or the rate at which each home should inject attack traffic into the network.

From the defense standpoint, monitors are typically installed at gateways to protect against DDoS attacks. Usually they employ packet inspection and keep track of different traffic features such as IP addresses and HTTP headers. We focus on a lightweight approach avoiding any information typically obtained from packet inspection as in [8] and rely only on packet (or byte) counts.

The system we study is shown in Figure 1. In the figure, the blue dotted arrows represent measurement data collected at home routers and sent periodically to a data fusion center for analysis. In fact, the data that we had access to and used in Section 5 was obtained by a measurement effort in which packet and

byte counts of upload/download traffic are collected at participant home-routers at minute intervals and sent to a fusion center. Blue arrows represent regular upload traffic, and the widths of the arrows indicate distinct traffic statistics. The botmaster can issue commands to the infected houses (in red) but, as shown in the figure, the attacker can choose not to use all the houses he controls to initiate an attack, to cause significant damage and yet remain covert.

2.2. Statistical Hypothesis Testing

Consider a collection of n homes where each, using its home-router, continuously measures upload traffic during a time slot, and sends this information to an ISP fusion center where detection takes place. We assume that there is an attacker who may or may not launch an attack during a time window.

The system administrator (henceforth known as admin) performs a hypothesis test on observations with the null hypothesis H_0 being that the attacker does not launch an attack and the alternate hypothesis H_1 that he does launch an attack. We are interested in the following question: *can the attacker launch an attack without being detected by admin and, if so, how large can such an attack be?*

Admin can tolerate some false positives, or cases when the statistical test incorrectly concludes an attack is under way. When correct, this rejection of H_0 is known as a false alarm, and, following standard nomenclature, we denote its probability by p_{FA}. Admin's test may also fail to indicate that an attack is taking place. Acceptance of H_0 when it is false is known as a missed detection, and we denote its probability by p_{MD}. Then, the sum $p_{FA} + p_{MD}$ characterizes the necessary tradeoff between false alarms and missed detections in the design of a hypothesis test.

Denote the upload traffic probability distribution in the absence of an attack (i.e. when H_0 is true) as $f_0(x)$, and in presence of attack (i.e. when H_1 is true) as $f_1(x)$. When $f_0(x)$ and $f_1(x)$ are known to admin, he can construct an optimal statistical hypothesis test (such as the Neyman-Pearson or likelihood

ratio test) that minimizes the *sum of error probabilities* [15, Ch. 13],

$$S_E := p_{FA} + p_{MD}. \tag{1}$$

2.3. Covertness

Next, we formally introduce the covertness criterion used throughout this work. This covertness criterion was proposed in the context of low probability of detection (LPD) communications in [10].

Definition 2.1. *An attack is covert provided that, for any $\epsilon > 0$, the attacker has a strategy for each n such that*

$$\liminf_n S_E \geq 1 - \epsilon. \tag{2}$$

The justification for this definition is the following. Assume that the optimal statistical hypothesis test that admin runs is such that $S_E > 1$. Then,

$$S_E = p_{FA} + p_{MD} = \mathbb{P}(\text{accept } H_1 \,|\, H_0 \text{ is true}) + \mathbb{P}(\text{accept } H_0 \,|\, H_1 \text{ is true}) > 1,$$

which yields

$$\mathbb{P}(\text{accept } H_1 \,|\, H_0 \text{ is true}) > \mathbb{P}(\text{accept } H_1 \,|\, H_1 \text{ is true})$$

and

$$\mathbb{P}(\text{accept } H_0 \,|\, H_1 \text{ is true}) > \mathbb{P}(\text{accept } H_0 \,|\, H_0 \text{ is true}).$$

Such a statistical hypothesis test cannot be optimal as if the attacker decreases the attack traffic the probability of errors calculated by admin should not increase. Hence, $S_E \in [0, 1]$, which shows that Definition 2.1 ensures that the maximal value of S_E should be reached for covertness, which agrees with the intuition.

Note that a sufficient condition for the attack not to be covert is if, for some $\epsilon \in (0, 1)$, there exists a detector such that

$$\limsup_n S_E < \epsilon. \tag{3}$$

According to Definition 2.1 a successful attacker must be covert for any target p_{FA}.

3. DDOS Model

We introduce our model to tackle the interplay between DDoS covert attacks and lightweight defences. As described in Section 2.1, we consider a population of n home-routers equipped with monitors, that periodically collect byte and packet counts of upload/download traffic that flows through each home-router. Time is divided into time slots (also called time windows) of duration of Δ seconds. At each time slot, one sample is collected from each home and transmitted to a server (fusion center). We start by considering the problem of determining the maximum damage that an attacker can cause without being detected.

Consider n observations collected during a time slot, where each observation corresponds to a different home-router. The models of regular and attack traffic at a given time slot are characterized by the following two random variables (rvs). Here $r = 1, \ldots, n$.

- X_r, the amount of regular traffic, measured in packets or bytes, uploaded from the r-th home;
- Y_r, the amount of attack traffic, measured in packets or bytes, uploaded from the r-th home.

Given an attack takes place, let χ_r be a rv that takes value 1 if home r is used by the botmaster in the attack (see Figure 1) and 0 otherwise, with

$$q(n) = \mathbb{P}(\chi_r = 1). \tag{4}$$

Let Z_r denote the amount of observed traffic at home-router r in a given time slot,

$$Z_r = \begin{cases} X_r & \text{if no attack occurs,} \\ X_r + \chi_r Y_r & \text{otherwise.} \end{cases}$$

Intuitively, if the attacker is too aggressive the probability of error by the admin detector will be zero. Alternatively, if the attacker is timid, he will not be detected, but the average total amount of data that he injects, $q(n) \sum_{r=1}^{n} \mathbb{E}[Y_r]$, will be limited. The objective of this paper is to quantity these intuitions.

To this end, we formulate DDoS detection as a statistical hypothesis testing problem where the null and alternative hypotheses are given as follows,

- H_0 (no attack taking place): $Z_r = X_r$,
- H_1 (attack taking place): $Z_r = X_r + \chi_r Y_r$.

Table 1 contains a glossary of notations used throughout this paper.

In order to derive achievability and converse results, we need to specify the distribution of the regular and attack traffic. For achievability results we assume that both traffic are modeled by Gaussian mixtures (see below). The Gaussian mixture model is motivated by an exploratory analysis of the dataset we had access to. A general traffic model will be allowed for the converse (see Section 4).

We now introduce both the regular and attack traffic models under which our achievability results will be obtained (cf. Theorems 4.1-4.2). We assume that the regular traffic X_r generated by home-router r in a time window is modeled by a mixture of I_r Gaussians with probability density function (pdf) given by

$$f_{0,r}(x) = \sum_{i=1}^{I_r} \frac{w_{0,i,r}}{\sqrt{2\pi\sigma_{0,i,r}^2}} e^{-\frac{(x-\mu_{0,i,r})^2}{2\sigma_{0,i,r}^2}}, \tag{5}$$

with

$$0 < w_{0,i,r} < 1, \quad \sum_{i=1}^{I_r} w_{0,i,r} = 1, \quad \sigma_{0,i,r} > 0, \quad i = 1, \ldots, I_r. \tag{6}$$

The pdf of the attack traffic Y_r at home-router r is independent of r and is given by the Gaussian mixture with pdf

$$v(x,n) = \sum_{j=1}^{J} \frac{w_{1,j}}{\sqrt{2\pi\sigma_{1,j}^2(n)}} e^{-\frac{(x-\mu_{1,j}(n))^2}{2\sigma_{1,j}^2(n)}}, \tag{7}$$

with $0 < w_{1,j} < 1$ for $j = 1, \ldots, J$ and $\sum_{j=1}^{J} w_{1,j} = 1$. We assume that $\inf_{n\geq 1} \sigma_{1,j}^2(n) > 0$ for $j = 1, \ldots, J$. Notice in (7) the (potential) dependency on n of the parameters of the attack traffic.

Denote by $\mu_{0,r} = \mathbb{E}[X_r]$ and $\sigma_{0,r}^2 = \mathrm{var}(X_r)$ the mean and variance of the regular traffic generated in a time window by home-router r, and by $\mu_1(n) =$

Table 1: Glossary of notations

Variable	Description
n	number of home-routers (compromised nodes)
$q(n)$	probability a home-router participates in the attack
Δ	time window (or slot) duration (= 1 minute)
$f_{0,r}(x)$	pdf under H_0 of traffic injected in a slot by home-router r
$v(x,n)$	pdf traffic injected in a slot by attacker
$f_{1,r}(x,n)$	pdf of regular and attack traffic injected in a slot by home-router r
$g_r(x,n)$	pdf under H_1 of regular and attack injected in a slot by home-router r
$\mu_{0,r}$	mean traffic injected in a slot by home-router r (packets)
$\sigma^2_{0,r}$	variance of traffic injected in a slot by home-router r
$\mu_1(n)$	mean traffic injected in a slot by attacker (packets)
$\sigma^2_1(n)$	variance of traffic injected in a slot by attacker
I_r	number of mixture components of traffic of home-router r, resp.
J	number of mixture components for attack traffic
$w_{0,i,r}$	mixture weight for component i of regular traffic of home r $(i = 1, \ldots, k_1)$
$w_{1,j}$	mixture weight for component j of attack traffic $(j = 1, \ldots, J)$
$\mu_{0,i,r}$	mean of component i of traffic of home-router r
$\sigma^2_{0,i,r}$	variance of component i of traffic of home-routeur r
$\mu_{1,j}(n)$	mean of component j of attack traffic
$\sigma^2_j(n)$	variance of component j of attack traffic
$\bar{z}$	average overall traffic in a given time window

$\mathbb{E}[Y_r]$ and $\sigma_1^2(n) = \text{var}(Y_r)$ the mean and variance of the attack traffic generated in a time window. Notice that the derivation of the theoretical results do not require the means $\mu_{0,r}$ $(r = 1, \ldots, n)$ and $\mu_1(n)$ to be nonnegative; in practice these means will be strictly positive (cf. Section 5). These quantities are given by $\mu_{0,r} = \sum_{i=1}^{I_r} w_{0,i,r}\mu_{0,i,r}$, $\sigma_{0,r}^2 = \sum_{i=1}^{I_r} w_{0,i,r}\sigma_{0,i,r}^2$, $\mu_1(n) = \sum_{j=1}^{J} w_{1,j}\mu_{1,j}(n)$, and $\sigma_1^2(n) = \sum_{j=1}^{J} w_{1,j}\sigma_{1,j}^2(n)$.

The sum of the regular and attack traffic $X_r + Y_r$ at home-router r when an attack takes place has pdf

$$f_{1,r}(x,n) = \sum_{i=1}^{I_r}\sum_{j=1}^{J} \frac{w_{0,i,r}w_{1,j}}{\sqrt{2\pi(\sigma_{0,i,r}^2 + \sigma_{1,j}^2(n))}} e^{-\frac{(x-\mu_{0,i,r}-\mu_{1,j}(n))^2}{2(\sigma_{0,i,r}^2+\sigma_{1,j}^2(n))}}. \tag{8}$$

Under H_0, the total traffic Z_r uploaded from home-router r has pdf $f_{0,r}(x)$ and under H_1 the pdf of the total traffic Z_r uploaded from home-router r has pdf $g_r(x,n)$ given by

$$g_r(x,n) = (1 - q(n))f_{0,r}(x) + q(n)f_{1,r}(x,n). \tag{9}$$

To avoid unnecessary complications we require that all parameters in $f_{0,r}(x)$ are uniformly bounded in r as $n \to \infty$, namely,

$$\text{(a)} \quad \max_{r\geq 1} I_r < \infty, \quad \text{(b)} \quad \sup_{r\geq 1} |\mu_{0,r}| < \infty, \quad \text{(c)} \quad 0 < \inf_{r\geq 1} \sigma_{0,r}^2 \leq \sup_{r\geq 1} \sigma_{0,r}^2 < \infty. \tag{10}$$

Conditions in (10) together with (6) imply

$$0 < \inf_{1\leq i\leq I_r, r\geq 1} w_{0,i,r}, \qquad \sup_{i=1,\ldots,I_r, r\geq 1} |\mu_{0,i,r}| < \infty \tag{11a}$$

$$0 < \inf_{i=1,\ldots,I_r r\geq 1} \sigma_{0,i,r}^2 \leq \sup_{i=1,\ldots,I_r, r\geq 1} \sigma_{0,i,r}^2 < \infty. \tag{11b}$$

Also under (10) (see Appendix F)

$$\sup_{r\geq 1} \mathbb{E}[|X_r - \mu_{0,r}|^3] < \infty. \tag{12}$$

In particular, conditions in (10)-(11) are satisfied if $f_{0,r} \in \{\psi_i, i = 1, \ldots, K\}$ for all $r \geq 1$, where ψ_i $(i = 1, \ldots, K,\ K < \infty)$ is the pdf of a Gaussian mixture or, equivalently if home-routers belong to K *different classes* in terms of the traffic

that they generate. Indeed, in this case all parameters of the Gaussian mixture in (5) take a finite number of (finite) values.

Denote by $f_0^{(n)}$ the joint density of the mutually independent random variables (rvs) $X_1, \ldots, X_n$ and by $g^{(n)}(x)$ the joint density of the mutually independent rvs $Z_1, \ldots, Z_n$. We have (cf. (5)-(8))

$$f_0^{(n)}(x) = \prod_{r=1}^{n} f_{0,r}(x_r, n), \quad g^{(n)}(x) = \prod_{r=1}^{n} g_r(x_r, n). \tag{13}$$

for $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

Last, we consider two settings. The first is when admin knows the parameters (i.e. pdf) of the attack model and the second is when these parameters are not known to admin. We present a classifier for each of these in Section 5.2, Type I for the case the attack model is completely known and Type II for the case it is not known. The Gaussian mixture representation of the network traffic in (5)-(7) is motivated and validated using real data in Section 5.

4. Theoretical Results

In this section we present our achievability and converse results.

4.1. Achievability

The first achievability result in Theorem 4.1 holds when both regular and attack traffic have Gaussian distributions. The second and more general achievability result in Theorem 4.2 holds when both regular and attack traffic are represented by mixtures of Gaussians. Both theorems exhibit square-root laws.

Theorem 4.1 (Achievability when home & attack traffic have Gaussian distributions)**.** *Assume that X_r and Y_r have Gaussian distributions, with mean $\mu_{0,r}$ and variance $\sigma^2_{0,r}$ for X_r and with mean $\mu_1(n)$ and variance $\sigma^2_1(n)$ for Y_r. We assume that admin knows the distribution of the attack traffic. The attack traffic is covert if*

$$\mu_1(n) = O(1), \quad 0 < \sup_n \sigma^2_1(n) < \inf_{r \geq 1} \sigma^2_{0,r}, \tag{14}$$

and

$$q(n)\mu_1(n) = \mathcal{O}(1/\sqrt{n}), \quad q(n)\sigma_1^2(n) = \mathcal{O}(1/\sqrt{n}). \tag{15}$$

Theorem 4.2 (Achievability when home & attack traffic are mixtures of Gaussians)**.** *Assume that the pdfs of X_r and Y_r are given in (5) and (7), respectively, and that admin knows the parameters of the attack traffic distribution. Under (10) the attack traffic is covert if (with $j = 1, \ldots, J$)*

$$\mu_{1,j}(n) = \mathcal{O}(1), \quad 0 < \sup_{1\le j\le J, n\ge 1} \sigma_{1,j}^2(n) < \inf_{1\le i\le I_r, r\ge 1} \sigma_{0,i,r}^2, \tag{16}$$

and

$$q(n)\mu_1(n) = \mathcal{O}(1/\sqrt{n}), \quad q(n)\sigma_1^2(n) = \mathcal{O}(1/\sqrt{n}). \tag{17}$$

When parameters of the attack traffic in (7) do not depend on n, Theorem 4.2 says that the attack is covert when

$$\max_{1\le j\le J} {\sigma_{1,j}}^2 < \inf_{1\le i\le I_r, r\ge 1} \sigma_{0,i,r}^2 \quad \text{and} \quad q(n) = \mathcal{O}(1/\sqrt{n}). \tag{18}$$

Theorems 4.1 and 4.2 imply that the total amount of traffic that an informed attacker can inject into the network grows as $O(\sqrt{n})$. As n grows, a covert attacker must inject less traffic per home, but the total amount of traffic is still unbounded as a function of the number of homes.

The proof of Theorem 4.1 is given in Appendix Appendix B. It uses the same argument as the proof of the more general result in Theorem 4.2 – given in Appendix Appendix C – but has the advantage of being much shorter. Before sketching out these proofs let us introduce some intermediary results.

Theorem 4.3 below relates the minimum – denoted by $S_E^\star$ – of the sum of error probabilities p_{FA} and p_{MD} to the total variance distance between pdfs $f_0^{(n)}$ and $g^{(n)}$. We recall that $f_0^{(n)}(x) = \prod_{r=1}^n f_{0,r}(x_r)$ is the joint pdf of $Z_1, \ldots, Z_n$ under H_0 and $g^{(n)}(x) = \prod_{r=1}^n g_r(x_r, n)$ is the joint pdf of $Z_1, \ldots, Z_n$ under H_1 for $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ – see Section 3.

Theorem 4.3. *[Theorem 13.1.1 in [15]]*

Using the observed values $\mathbf{z}_n := (z_1, \ldots, z_n)$ of $Z_1, \ldots, Z_n$, any test accepting H_0 if $f_0^{(n)}(\mathbf{z}_n) < g^{(n)}(\mathbf{z}_n)$ and rejecting H_0 if $f_0^{(n)}(\mathbf{z}_n) > g^{(n)}(\mathbf{z}_n)$ minimizes S_E.

Furthermore, the minimum S_E is given by

$$S_E^\star = 1 - T_V\left(f_0^{(n)}, g^{(n)}\right),$$

where $T_V(u,v) := \int |u(x) - v(x)|dx$ is the total variance distance between pdfs u and v.

Following [10] we say that an attack is ϵ-covert ($0 < \epsilon < 1$) if $\liminf_n S_E^\star \geq 1 - \epsilon$, or equivalently by Theorem 4.3, if $\limsup_n T_V\left(f_0^{(n)}, g^{(n)}\right) \leq \epsilon$. Calculating $T_V\left(u^{(n)}, v^{(n)}\right)$ is usually very difficult and our problem is no exception. Instead of working directly with $T_V\left(f_0^{(n)}, g^{(n)}\right)$ we use the upper bound reported in the lemma below.

Lemma 4.1 (Upper bound on total variation distance)**.**
For all $n \geq 1$,

$$T_V\left(f_0^{(n)}, g^{(n)}\right) \leq \frac{1}{2}\sqrt{\prod_{r=1}^{n}(1 + q(n)^2 C_r(n)) - 1}, \tag{19}$$

where the nonnegative constant $C_r(n)$, known as the Fisher information constant [16], is defined by

$$C_r(n) = -1 + \int_{\mathbb{R}} \frac{f_{1,r}(x,n)}{f_{0,r}(x)}dx. \tag{20}$$

The proof of Lemma 4.1 is given in Appendix Appendix A.

Corollary 4.1. *Fix $\epsilon > 0$. If $\sum_{r=1}^{n} \log\left(1 + q(n)^2 C_r(n)\right) = \mathcal{O}(1)$ then*

$$\limsup_n T_V\left(f_0^{(n)}, g^{(n)}\right) \leq \epsilon, \tag{21}$$

in which case the attack is covert by the definition of covertness in (2.1) and Theorem 4.3.

Sketch of the proofs of Theorems 4.1 and 4.2: In Theorem 4.1 the Fisher constant $C_r(n)$ defined in (20) is given by (see (B.2))

$$C_r(n) = -1 + \sigma_{0,r}^2 e^{\frac{\mu_1^2(n)}{\sigma_{0,r}^2 - \sigma_1^2(n)}} / \sqrt{\sigma_{0,r}^4 - \sigma_1^4(n)}.$$

From this expression we show that $nq^2(n)\sup_{1\leq r\leq n} C_r(n) = \mathcal{O}(1)$ under (14)-(15) and the proof of Theorem 4.1 follows by invoking Corollary 4.1. The proof of Theorem 4.2 is similar but the more complicated value obtained for $C_r(n)$ in (C.1) makes it more tedious.

4.2. Converse

Denote by $\bar{z}$ the average traffic collected from the n homes in a given time slot. Given that $\bar{z}$ increases in the face of attacks, the test has the following threshold structure,

$$\bar{z} \gtrless_{H_0}^{H_1} \tau, \tag{22}$$

where τ is a threshold to determine if there is an attack in the network based on $\bar{z}$.

The converse theorem below holds for *any probability distribution* of the regular traffic X_r and of the attack traffic Y_r.

Theorem 4.4 (Converse, general distributions)**.**

Assume arbitrary probability distributions for the mutually independent rvs $X_1, \ldots X_n$, with mean $\mu_{0,r}$ and strictly positive variance $\sigma^2_{0,r}$ for X_r. We assume that $Y_1, \ldots, Y_n$ are independent and identically distributed rvs with mean $\mu_1(n)$ and variance $\sigma_1^2(n)$. We assume that admin does not know the attack distribution. We further assume that

$$\sup_{r\geq 1} |\mu_{0,r}| < \infty,\ 0 < \inf_{r\geq 1} \sigma^2_{0,r} \leq \sup_{r\geq 1} \sigma^2_{0,r} < \infty,\ \sup_{r\geq 1} \mathbb{E}[|X_r - \mu_{0,r}|^3] < \infty. \tag{23}$$

The attacker is not covert if [1]

$$\lim_n \sqrt{n} q(n)\mu_1(n) = +\infty,\ var(\chi_r Y_r) = q(n)\left(\sigma_1^2(n) + (1-q(n))\mu_1^2(n)\right) = \mathcal{O}(1). \tag{24}$$

Let us specialize Theorem 4.4 to the case where X_r and Y_r have pdfs given in (5) and (7) and home-routers belong to a finite number of classes. In this

[1] First condition in (24) can be replaced by $q(n)\mu_1(n) = \omega(1/\sqrt{n})$ if $\liminf_{n\geq 1} \mu_1(n) > 0$.

case, conditions (23) are satisfied (see discussion in Section 3) and Theorem 4.4 becomes,

Corollary 4.2 (Converse, mixture of Gaussian distributions)**.**

Assume that X_r and Y_r have pdfs given in (5) and (7) and that home-routers belong to a finite number of classes. Assume admin does not know the parameters of $v(x,n)$ in (7). The attacker is not covert if

$$\lim_n \sqrt{n} q(n)\mu_1(n) = +\infty,\ var(\chi_r Y_r) = q(n)\left(\sigma_1^2(n) + (1-q(n))\mu_1^2(n)\right) = \mathcal{O}(1). \tag{25}$$

The proof of Theorem 4.4 is given in Appendix Appendix G. Let us briefly discuss it. It consists of finding an upper bound on the error experienced by the classifier implementing threshold policy (22). To that aim, we first determine a threshold τ such that the corresponding probability of false alarm, p_{FA}, is upper bounded by a constant α. Then, given τ we show that the probability of miss-detection, p_{MD}, can be made arbitrarily close to 0 as n grows to infinity provided conditions in (25) hold. As the latter holds for any value of α, together those two bounds imply that the sum of error probabilities S_E can be made arbitrarily close to 0 as n grows.

Note that the proof of tightness is constructive, in the sense that it follows, in essence, the methodology to parametrize a Neyman-Pearson classifier [17]. The threshold selected by the Neyman-Pearson classifier is typically chosen to satisfy a constraint on the probability of false alarms, noting that the probability of miss-detection is minimized. As indicated above, we show that if conditions in (25) hold such a minimum can be made arbitrarily close to 0 as n grows to infinity, for any given upper bound on the probability of false alarm.

Theorems 4.2 and 4.4 highlight a phase transition at $1/\sqrt{n}$ which is apparent through the first conditions in (17) and (25), namely, the attack is covert if the expected attack traffic injected at a home-router in a slot behaves as $1/\sqrt{n}$ when n is large whereas it is not convert if this quantity decreases to zero not as fast as $1/\sqrt{n}$; note, however, that the variance of the attack traffic also plays a role (see second conditions in (17) and (25)) in the transition 'covert - not covert'.

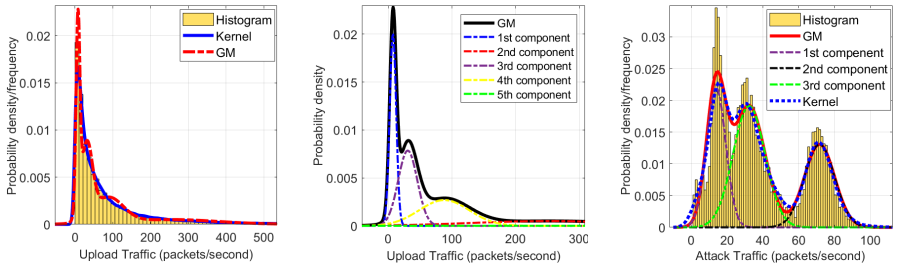
Remark 1. *If admin knows the attack distribution it can implement a test such that H_0 is accepted if $f_0^{(n)}(\cdot) < g^{(n)}(\cdot)$ and H_0 is rejected if $f_0^{(n)}(\cdot) > g^{(n)}(\cdot)$ so that Theorem 4.3 can be used to prove Theorems 4.1- 4.2. However, it should be clear that these theorems keep holding when admin does not know the parameters of attack traffic distribution as it cannot perform better with less knowledge.*

Theorem 4.4 (and Corollary 4.2) has been obtained for the test in (22), a test that does not use information on the attack traffic distribution. We used this test for its simplicity (recall that to establish a converse result it is enough to exhibit a test yielding non covertness). However, it should also be clear that Theorem 4.4 keeps holding when admin knows the attack traffic distribution as it cannot perform less effectively with more knowledge.

5. Evaluation

Results obtained in Section 4 are asymptotic results, holding when n, the number of homes, goes to infinity. It is therefore interesting to investigate the "robustness" of these results when n is finite as is always the case in practice. To do so, we have relied on real data for the regular traffic and synthetic data for the attack traffic. Real data was collected by a mid-sized ISP network, with whom we partnered to collect traffic from home-routers.[2] To carry out this program, namely check the validity of Theorems 4.1, 4.2, and 4.4 when n is finite, we first need to identify the regular and attack traffic distributions (Section 5.1) and build detectors (Section 5.2) enabling admin to detect whether or not an attack has taken place. Two detectors are considered: one that uses knowledge about the distribution of the attack traffic and another where such knowledge is not required. For finite values of n, S_E sharply changes as a function of either the fraction of homes used by the attacker or the amount of injected traffic (Section 5.3). These results are in agreement with the square root law we discovered, allowing us to assess the minimum population size required to reach asymptotic

[2]To preserve anonymity, we will disclose the ISP in the final version of the paper.



(a) (b) (c)

Figure 2: (a) Histogram of upload traffic (packet per second) measured in one minute slot, kernel and Gaussian mixture models fit it; (b) five components of that mixture model; (c) histogram of attack traffic, models, and three Gaussian components fit it.

results.

5.1. Regular and Attack Traffic Distributions

We use data collected from network interfaces of more than 5000 home-routers. The selected home-routers were equipped with monitoring software to conduct a data collection campaign at home gateways. These routers gather information about network usage. For the purpose of this work, we use packet counts. The measurements correspond to the traffics uploaded by every user at all *one-minute time slots* between March 1st 2020 and April 30th 2020 (inclusive). Figure 2(a) shows the histogram of measurements and how Gaussian mixture distribution fits to it. Although the theory holds for the general case where traffic from homes have different distributions, to facilitate our experiments we assume that the home traffic distributions are identical (i.e. $f_{0,r} = f_{0,r'}$ for all $r, r' \geq 1$).

We used the EM (Expectation-Maximization) algorithm [18] to fit data to a Gaussian mixture model. Assessing goodness-of-fits using Kolmogorov-Smirnov (KS) test [19], we observed that traffic data can be characterized by a mixture of five Gaussian distributions (Figure 2(b)). Table 2 presents the estimates of model parameters for each component of this mixture.

In addition to measurements collected from the ISP network, we used a

Table 2: Weights and parameters of Gaussian mixture model fitted to regular traffic data (pkts/sec).

i	$w_{0,i}$	$\mu_{0,i}$	$\sigma_{0,i}$
1	0.26	7.69	5.09
2	0.15	253.19	122.75
3	0.30	30.90	14.97
4	0.26	88.09	39.37
5	0.03	1017.14	737.28

dataset for attack traffic generated by controlled experiments done in [8] using real Mirai code and estimate the distribution of traffic generated by a typical DDoS attack. The Mirai attacks use default parameters from one of its publicly available source codes.[3] Figure 2(c) shows the histogram of attack traffic in packets, where a mixture of three Gaussians provides an excellent fit to it.

Motivated by the aforementioned attack in a controlled environment, we generate attack traffic from a Gaussian mixture distribution with three components, where

$$\mu_{1,i}(n) = \delta c_{1,i} n^{-\alpha} \text{ and } \sigma_{1,i}(n)^2 = \delta c_{2,i} n^{-\alpha}, \quad i = 1, 2, 3. \tag{26}$$

Estimates of the other parameters are reported in Table 3.

When the attacker uses all homes ($q(n) = 1$), it sends on average a total amount of $60 \times \delta \sum_{i=1}^{3} w_{1,i} c_{1,i} n^{1-\alpha}$ packets from the n homes per slot (recall that each slot corresponds to 60 seconds). When considering an attacker that issues an attack from a fraction of the homes, we let

$$q(n) = c_q n^{-\beta}. \tag{27}$$

Parameters α and β will vary according to our experimental goals.

It is worth mentioning that the regular traffic considered in our evaluation is obtained directly from our real traces. We used Gaussian mixture models

[3] `https://github.com/jgamblin/Mirai-Source-Code/pull/38`

Table 3: Parameters of Gaussian mixture model fitted to attack traffic data (pkts/sec).

i	$w_{1,i}$	$c_{1,i}$	$c_{2,i}$
1	0.29	13.91	5.35
2	0.28	71.64	8.42
3	0.43	32.29	8.92

only to characterize attack traffic generated by Mirai (Eq. (26)) and to compute the likelihood ratios needed to parametrize the first detector introduced in the sequel.

5.2. Attack Identification

Once the data has been collected, we need a detector[4] to decide whether or not an attack has taken place. We will consider two detectors, one which uses the attack traffic distribution (Type I) and another one which does not (Type II).

Introduce the likelihood ratio

$$\Lambda(z_1,\ldots,z_n)=\begin{cases}\frac{\prod_{r=1}^{n} f_{0,r}(z_r)}{\prod_{r=1}^{n} f_{1,r}(z_r,n)}, & q(n)=1,\\ \frac{\prod_{r=1}^{n} f_{0,r}(z_r)}{\prod_{i=1}^{n} g_r(z_r,n)}, & q(n)\in(0,1),\end{cases} \tag{28}$$

where pdfs $f_{0,r}(x)$, $f_{1,r}(x,n)$, and $g_r(x,n)$ are defined in (5), (8), and (9), respectively. The first case (i.e., $q(n)=1$) accounts for when the attacker launches an attack from all homes and the second case (i.e., $q(n)\in(0,1)$) is when it chooses a fraction of homes to launch an attack from. In (28) z_r is one realization of the rv Z_r (see Section 3), the amount of traffic measured at home-router r in one slot.

The Type I detector is given by the threshold policy

$$\Lambda(z_1,\ldots,z_n)\gtrless_{H_1}^{H_0}\tau, \tag{29}$$

[4]Referred to as a test in the theoretical part of this work as is usually the case in hypothesis testing theory.

where τ is a threshold set to satisfy a desired probability of false alarm, p_{FA}. This detector is optimal from Neyman and Pearson Lemma [20, p. 491]. In the following $\Lambda(z_1, \ldots, z_n)$ is denoted by Λ.

Algorithm 1: Computes τ when $p_{FA} = \zeta$.

Input : ζ (prefined value of p_{FA}), number of homes n, counter M

Output: threshold τ

1 **for** $i = 1 \rightarrow M$ **do**

2 | generate $z_1, \ldots, z_n$ from real data (i.e. regular traffic only);

3 | compute likelihood ratio Λ;

4 | list(i) = Λ;

5 **end**

6 τ = the $\lfloor \zeta M \rfloor$-th smallest value of list.

Given p_{FA} is set to a predefined value ζ, the threshold τ is computed from the Monte Carlo algorithm described in Algorithm 1 below. First, we generate one sample of the traffic from the *real data* (line 2). Then, we compute the corresponding likelihood ratio Λ from (28) (line 3). This process is repeated M times and the collection of likelihood ratios is sorted in ascending order, $\Lambda_{(1)} \leq \ldots \leq \Lambda_{(M)}$. This means that $M_0 := \lfloor \zeta M \rfloor$ values of Λ have been incorrectly classified as an attack. The threshold τ is then set to be $\Lambda_{(M_0)}$.

Once the threshold τ has been computed, we estimate the probability of miss detection, p_{MD}, also via a Monte Carlo iterative approach. The synthetic attack traffic (see Section 5.1) is added to the regular traffic. At each iteration, one sample $z_1, \ldots, z_n$ is collected from the data and Λ is calculated from (28). This process is repeated L times. We then calculate the number of Λ's that are larger than τ, say L_0, and p_{MD} is obtained as $p_{MD} = L_0/L$.

The Type II detector does not have any knowledge on the attack traffic. We will however assume that when admin uses it it will know if conditions in (23) are satisfied. As already mentioned in Section 4 we will work with the Type II

detector given by

$$\bar{z}_n \gtrless_{H_0}^{H_1} \tau, \tag{30}$$

with $\bar{z}_n := \frac{1}{n}\sum_{r=1}^{n} z_r$. Note that the null and alternative hypotheses subsumed by this Type II detector corresponds to $H_0 : \mu = \mu_0(n)$ and $H_1 : \mu > \mu_0(n)$, respectively. The threshold τ when the probability of false alarm is set is to ζ can be determined by using Algorithm 1 upon replacing Λ by $\bar{z}_n$, and from there one obtains p_{MD} as explained for the Type I detector.

5.3. Phase Transition and Square Root Law

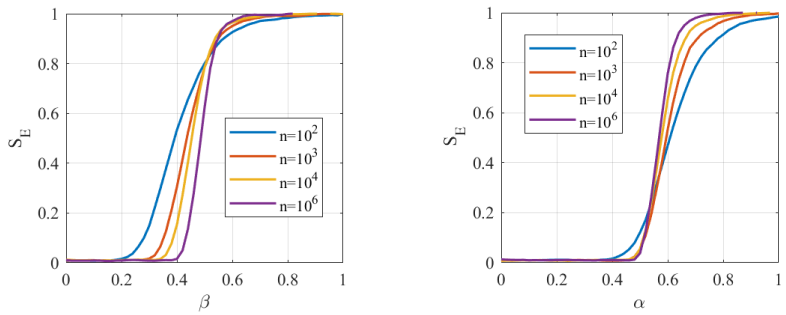
This section focuses on phase transition (transition from a regime wherein the attacker is detected with high probability to a regime wherein the attacker cannot be detected) and its square root law companion identified in the theoretical results in Section 4.

Figure 3(a) displays S_E, the sum of p_{FA} and p_{MD}, when $p_{FA} = 0.01$, as a function of the fraction of infected homes used by the attacker, $q(n) = n^{-\beta}$. We let β vary from 0 to 1, $\delta = 4$ and $\alpha = 0$ so that the rate at which each active home injects attack traffic does not depend on n (see (26)).

Figure 3(b) displays S_E with $p_{FA} = 0.01$ and $\delta = 2$, this time as a function of the total average attack traffic $\mu_1(n) = 60 \times 2 \times \sum_{i=1}^{3} w_{1,i} c_{1,i} n^{1-\alpha}$ injected in a slot when α varies from 0 to 1 and when all homes are used by the attacker ($q(n) = 1$).

Plots in each figure correspond to different values of n, the number of homes, with $n \in \{10^2, 10^3, 10^4, 10^6\}$. Both figures have been obtained with the Type I detector in (29).

To interpret the results we need to have in mind that very small values of S_E imply that the attack will be detected, otherwise the attack will be undetected. The first observation is that all curves exhibit a "phase transition", which becomes sharper as n increases. For $n \geq 10^3$ the transition occurs in Figure 3(b) around $\alpha = 0.5$ which corresponds to $\mu_1(n)$ being of the order of $1/\sqrt{n}$, in agreement with the theory. For $n = 10^6$ the transition occurs in Figure 3(a)

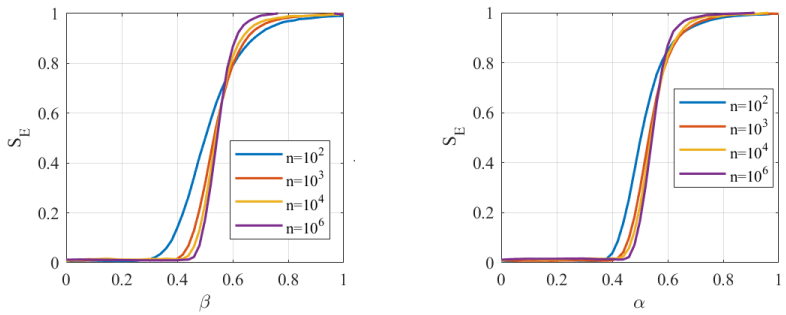


(a) (b)

Figure 3: Phase transition analysis for Type I detectors, with $p_{FA} = 0.01$. (a): fraction of homes used by attacker given by $q(n) = n^{-\beta}$, $\alpha = 0$, $\delta = 4$; as β grows, total attack traffic decreases and the sum of probabilities of errors (S_E) transitions from 0 to 1. For $n = 10^6$ phase transition occurs around $\beta = 0.5$, in agreement with square root law. (b): all homes used by attacker ($q(n) = 1$), average total traffic injected by the attacker given by $60 \times \delta \sum_{i=1}^{3} w_{1,i} c_{1,i} n^{1-\alpha}$, where $\delta = 2$; as α grows total attack traffic decreases, and S_E transitions from 0 to 1. For $n \geq 10^3$, sharp phase transition occurs around $\alpha = 0.5$, in agreement with square root law.

around $\beta = 0.5$ corresponding to $q(n) = 1/\sqrt{n}$, again in agreement with the theory.

Figure 4 reports results on the square root law and phase transitions accounting for the Type II detector, in a reference scenario with $\delta = 20$, $c_q = 1$, and $p_{FA} = 0.01$. Figure 4(a) shows S_E as a function of β, letting $\alpha = 0$. As β increases the total attack traffic decreases and S_E transitions from 0 to 1. The larger the number of homes, the sharper the transition. When $n = 1000$ the phase transition is already noticeable, and when $n = 10^6$ the sharp transition occurs at $\alpha = 0.5$, which corresponds to $q(n) = 1/\sqrt{n}$, in agreement with the square root law. Figure 4(b) shows the error probability accounting for an attacker that issues the attack from all homes ($\beta = 0$ and $q(n) = 1$), and controls the rate at which traffic is injected from each home. The average total traffic injected by the attacker is given by $60 \times \delta \sum_{i=1}^{3} w_{1,i} c_{1,i} n^{1-\alpha}$, where $\delta = 20$. As α increases, the attacker becomes less aggressive, and S_E increases. When $n = 10^3$ we already observe a sharp transition in the sum of probabilities of



(a) (b)

Figure 4: Phase transition analysis for Type II detectors, with $p_{FA} = 0.01$. (a): fraction of homes used by attacker given by $q(n) = n^{-\beta}$, $\alpha = 0$, $\delta = 20$; as β grows, total attack traffic decreases and the sum of probabilities of errors (S_E) transitions from 0 to 1. For $n = 10^3$ phase transition occurs around $\beta = 0.5$, in agreement with square root law. (b): all homes used by attacker ($q(n) = 1$), average total traffic injected by the attacker given by $60 \times \delta \sum_{i=1}^{3} w_{1,i} c_{1,i} n^{1-\alpha}$, where $\delta = 20$; as α grows total attack traffic decreases, and S_E transitions from 0 to 1. For $n \geq 10^3$, sharp phase transition occurs around $\alpha = 0.5$, in agreement with square root law.

errors (S_E) at $\alpha = 0.5$, which corresponds to $\mu_1(n)$ being of the order of $1/\sqrt{n}$, again in agreement with the theory.

As Type I detectors make use of more information, the phase transition occurs for small values of n. Indeed, for $\delta \leq 4$ the Type I detectors shown in Figure 3 already exhibit asymptotic behavior for values of n greater than 10^4. Type II detectors, in contrast, require a larger number of homes to reach asymptotic behavior in that setting (not shown in the paper). For this reason, in Figure 4 we set $\delta = 20$, which corresponds to a more aggressive attacker, evidencing the phase transition for $n \geq 10^3$.

Next, we consider an attacker that can jointly control the attack rate per home and the fraction of active homes. The attacker controls the rate at which each home injects traffic into the network and the fraction of homes issuing an attack through parameters α and β (Eq. (26) and (27)), respectively. Noting that the average amount of traffic injected into the network during an attack is proportional to $n^{1-\alpha-\beta}$, Figure 5 shows the error probability as a function of α

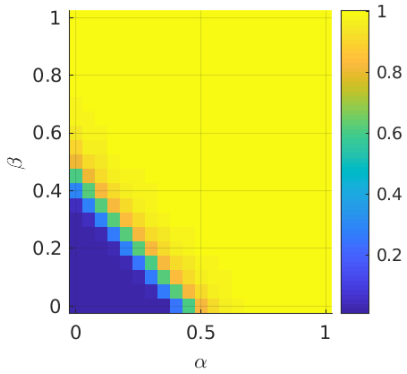


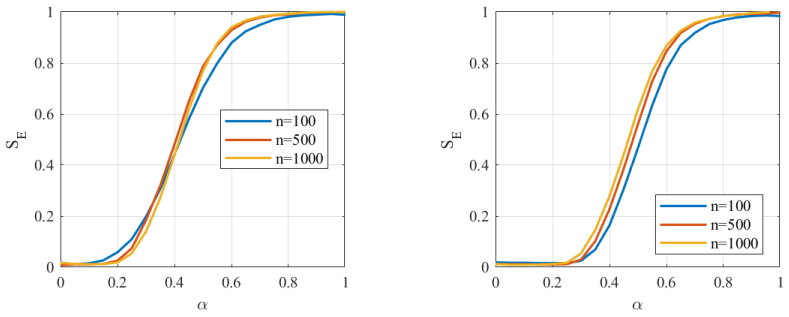
Figure 5: Joint control of fraction of active homes and rate per home, through α and β, respectively, (i.e., $q(n) = c_q n^{-\beta}$, $\mu_{1,i} \propto \delta n^{-\alpha}$, and $\sigma_{1,i} \propto \delta n^{-\alpha}$), accounting for Type II detectors, with $p_{FA} = 0.01$, $c_q = 1$, $\delta = 20$, and $n = 10^3$. There is a phase transition close to the line where $\alpha + \beta = 0.5$, in agreement with the theory.

and β. It indicates a phase transition close to the line where $\alpha + \beta = 0.5$. In light of Type II detectors, the effect of α and β on the error probabilities occurs through their sum $\alpha + \beta$, in agreement with the theory.

Finally, we examine a more realistic scenario where in each iteration of Algorithm 1 (lines 1-5), the generated regular traffics $z_1, \ldots, z_n$ are restricted to the real data of a random single time slot. Figures 6(a) and 6(b) show S_E as a function of α accounting for the Type I detector with $\delta = 1$ and Type II detector with $\delta = 20$, respectively, where $\beta = 0$, $c_q = 1$, and $p_{FA} = 0.01$. We observe that the phase transition is evident at $\alpha = 0.5$, in agreement with the square root law, although the distribution of the data during each single time slot is not necessarily the same as the distribution of all data.

6. Related Work

In this section, we briefly review related literature on the three main topics pertaining this work: volume-based DDoS attacks, covertness and statistical hypothesis testing.



(a) (b)

Figure 6: Phase transition analysis for (a) Type I detector with $\delta = 1$ and (b) Type II detector with $\delta = 20$, where regular traffics correspond to the real data of random single time slots and the attacker selects all homes.

6.1. DDoS Detection

The prevalence of volume-based DDoS attacks motivates a vast literature on their properties [21, 22] and early detection [23, 24, 25]. Fundamental limits of DDoS attacks have been investigated by Fu and Modiano [21] assuming a graph topology comprising attackers and servers behind a load balancer. The authors establish conditions under which attackers can make the network unstable, without accounting for covertness.

Machine learning is typically used for DDoS detection, e.g., for feature selection [26] or to leverage specific aspects of control protocols used by botnets [27, 28]. Most of the works in this line of research consider deep packet inspection as a viable alternative, a notable exception being [8]. In [8] the authors consider lightweight strategies for attack detection solely based on statistics of byte counts and packet counts. In this paper we focus on such privacy-preserving methods that would still work with encrypted traffic, and study the covertness of attackers against volume-based detectors.

6.2. Covertness

Our work is related to recent work on low probability of detection (LPD) communications, which has been mostly studied in the realm of wireless com-

munications [11, 12, 13, 10]. The LPD problem focuses on determining the maximum amount of information that a party, Alice, can reliably transmit to a receiver, Bob, subject to a constraint on the detection probability by a warden, Willy [14].

Bash et al. [10] shows that LPD communication on wireless Gaussian channels yields a square root law. Although the square root law found in the wireless setting is similar in spirit to the one derived in the present work, our work differs from [10] in at least two aspects. In the communications setting, the assumption is that Willy has complete information. The warden can design a classifier with complete information about Alice's power, which corresponds to our Type I classifier. We account for a second type of classifier, that has no information regarding the attack traffic other than that it is Gaussian Mixture distributed. Second, our application is DDoS attacks and determining how much traffic an attacker can inject into the network, whereas [10] considers users interested in communicating through a wireless channel. In addition, our theoretical results are evaluated and validated using real ISP traffic which is out of the scope of [10].

In [29], we track a different setting where the admin can leverage several network traffic features to improve detection accuracy. Then, under the assumption that the joint distribution of those features is multivariate Gaussian, we show that an DDoS attack is covert if its corresponding traffic features scale according to the square root of the number of compromised homes.

6.3. Statistical Hypothesis Testing

Statistical hypothesis testing and its extensions, including sequential hypothesis testing and sequential probability ratio tests (SPRT), are the pilars of statistical inference. SPRT has found its applications in the security field for port scan detection [30] and detection of attacks in mobile wireless networks [31, 32], to name a few.

One of the key ingredients of statistical hypothesis tests is a notion of distance between distributions, e.g., to determine if the normal traffic distribution

is "close" to a given sample. The literature on detection systems based on statistical hypothesis testing [33, 34, 35, 36, 37, 38] encompasses many notions of distance between distributions, including those based on Hartley entropy, Shannon entropy, Renyi entropy [39], and its variations [40], as well as KL divergence [41] and other measures of information gap between distributions. In this paper, we rely on the total variation distance, noting that future work involves considering other measures to assess distance between distributions, e.g., to derive non-asymptotic bounds.

In its simplest form, statistical hypothesis testing involves the characterization of the normal behavior of a system through a statistical model, followed by statistical tests to determine if the unknown samples are well captured by the model. In [30], for instance, the authors characterize port scans using random walks. The detection of attacks is based on determining if an observed random walk across the ports of a system can be well described by one of two stochastic processes, corresponding to malicious or authorized remote hosts scanning the network. In our work, we focus on the willingness of the attacker to remain covert, reporting results that are complementary to [31, 32, 30].

7. Conclusion

Botnets have reached impressive sizes counting with thousands of compromised nodes. Although an early detection of malicious traffic from those nodes can potentially prevent them from producing spectacular attacks, the fundamental limits on the accuracy of traffic classifiers must be taken into account when assessing their potential benefits. In this paper we established fundamental laws on the amount of traffic that an attacker can inject into a network, as a function of the size of the botnet, while still remaining covert. In particular, we show that in a scenario where all traffic is encrypted, and volume-based detectors are the sole viable solution, the amount of covert traffic can grow as the square root of the size of the botnet. Through numerical experiments parametrized with traffic collected from a mid-sized ISP, we have indicated that

the established laws capture the behavior of the considered classifiers in realistic settings, paving the way towards a foundational understanding of the intrinsic DDoS attack regimes.

Acknowledgments

This work was partially supported by grants from CNPq, CAPES, FAPERJ and by international cooperative grants from MCTIC-FAPESP, NSF EAGER 1740895/MCTIC-RNP and Army Research Labs (ARL) W911NF-17-2-0196.

References

[1] A. Marzano, D. Alexander, O. Fonseca, E. Fazzion, C. Hoepers, K. Steding-Jessen, M. H. Chaves, Í. Cunha, D. Guedes, W. Meira Jr., The evolution of Bashlite and Mirai IoT botnets, in: 2018 IEEE Symposium on Computers and Communications (ISCC), Natal, Brazil, 2018, pp. 00813–00818. doi:https://doi.org/10.1109/ISCC.2018.8538636.

[2] A. Ramachandran, D. Dagon, N. Feamster, Can DNS-based blacklists keep up with bots?, in: CEAS 2016 – The Third Conference on Email and Anti-Spam, Mountain View, California, USA, 2006.

[3] R. Doshi, N. Apthorpe, N. Feamster, Machine learning ddos detection for consumer internet of things devices, in: 2018 IEEE Security and Privacy Workshops (SPW), IEEE, 2018, pp. 29–35.

[4] A. Networks, Q4 2019 - The State of DDoS Weapons Report, 2019, `https://www.a10networks.com/marketing-comms/reports/state-ddos-weapons/` (2019).

[5] E. U. A. for Cybersecurity, Distributed Denial of Service:January 2019 - April 2020, `https://www.enisa.europa.eu/topics/threat-risk-management/threats-and-trends/etl-review-folder/etl-2020-denial-of-service` (2020).

[6] Cisco, Cisco Annual Internet Report (2018-2023) White Paper, https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html (2020).

[7] Cloudflare, Network-layer DDoS attack trends for Q2 2020 (2020).

[8] G. Mendonça, G. H. A. Santos, E. de Souza e Silva, R. M. M. Leao, D. S. Menasche, D. Towsley, An Extremely Lightweight Approach for DDoS Detection at Home Gateways, in: 2019 IEEE International Conference on Big Data, 2019, pp. 5012–5021. doi:10.1109/BigData47090.2019.9006548.

[9] H. Haddadi, V. Christophides, R. Teixeira, K. Cho, S. Suzuki, A. Perrig, Siotome: An edge-ISP collaborative architecture for IoT security, Proc. IoTSec (2018).

[10] B. A. Bash, D. Goeckel, D. Towsley, Square root law for communication with low probability of detection on AWGN channels, in: 2012 IEEE International Symposium on Information Theory Proceedings, IEEE, 2012, pp. 448–452.

[11] R. Soltani, D. Goeckel, D. Towsley, A. Houmansadr, Fundamental limits of covert packet insertion, IEEE Transactions on Communications (2020).

[12] K.-W. Huang, H.-M. Wang, D. Towsley, H. V. Poor, LPD communication: A sequential change-point detection perspective, IEEE Transactions on Communications (2020).

[13] B. Jiang, P. Nain, D. Towsley, Covert cycle stealing in a single FIFO server, ACM Transactions on Modeling and Performance Evaluation of Computing Systems (ToMPECS)ArXiv preprint arXiv:2003.05135 (2021. To appear).

[14] S. Yan, X. Zhou, J. Hu, S. V. Hanly, Low probability of detection communication: Opportunities and challenges, IEEE Wireless Communications 26 (5) (2019) 19–25.

[15] E. L. Lehmann, J. P. Romano, Testing Statistical Hypotheses, Springer Science & Business Media, 2006 (3rd edition).

[16] S. Kullback, Information Theory and Statistics, Dover Publications, 1968.

[17] C. Scott, R. Nowak, A Neyman-Pearson approach to statistical learning, IEEE Transactions on Information Theory 51 (11) (2005) 3806–3819.

[18] T. K. Moon, The expectation-maximization algorithm, IEEE Signal Processing Magazine 13 (6) (1996) 47–60.

[19] F. J. Massey Jr., The Kolmogorov-Smirnov test for goodness of fit, Journal of the American statistical Association 46 (253) (1951) 68–78.

[20] D. P. Bertsekas, J. N. Tsitsiklis, Introduction to Probability, Athena Scientific, 2008.

[21] X. Fu, E. Modiano, Fundamental limits of volume-based network dos attacks, Proceedings of the ACM on Measurement and Analysis of Computing Systems 3 (3) (2019) 1–36.

[22] J. Mirkovic, P. Reiher, A taxonomy of ddos attack and ddos defense mechanisms, ACM SIGCOMM Computer Communication Review 34 (2) (2004) 39–53.

[23] T. Andrysiak, L. Saganowski, DDoS Attacks Detection by Means of Statistical Models, in: Proceedings of the 9th International Conference on Computer Recognition Systems CORES 2015, Springer, 2016, pp. 797–806.

[24] Y. Tao, S. Yu, DDoS attack detection at local area networks using information theoretical metrics, in: 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, 2013, pp. 233–240.

[25] X. Ma, Y. Chen, DDoS detection method based on chaos analysis of network traffic entropy, IEEE Communications Letters 18 (1) (2013) 114–117.

[26] R. Doshi, N. Apthorpe, N. Feamster, Machine Learning DDoS Detection for Consumer Internet of Things Devices, in: 2018 IEEE Security and Privacy Workshops (SPW), 2018, pp. 29–35, iSSN: null.

[27] M. Ozçelik, N. Chalabianloo, G. Gur, Software-Defined Edge Defense Against IoT-Based DDoS, in: 2017 IEEE International Conference on Computer and Information Technology (CIT), 2017, pp. 308–313.

[28] Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, A. Shabtai, D. Breitenbacher, Y. Elovici, N-baIoT: Network-based detection of IoT botnet attacks using deep autoencoders, IEEE Pervasive Computing 17 (3) (2018) 12–22.

[29] A. R. Ramtin, P. Nain, D. Towsley, E. de Souza e Silva, D. S. Menasche, Are covert ddos attacks facing multi-feature detectors feasible?, ACM SIGMETRICS Performance Evaluation Review (2021).

[30] J. Jung, V. Paxson, A. W. Berger, H. Balakrishnan, Fast portscan detection using sequential hypothesis testing, in: IEEE Symposium on Security and Privacy, 2004. Proceedings. 2004, IEEE, 2004, pp. 211–225.

[31] J.-W. Ho, M. Wright, S. K. Das, Fast detection of mobile replica node attacks in wireless sensor networks using sequential hypothesis testing, IEEE transactions on mobile computing 10 (6) (2011) 767–782.

[32] J.-W. Ho, M. Wright, S. K. Das, ZoneTrust: Fast zone-based node compromise detection and revocation in wireless sensor networks using sequential hypothesis testing, IEEE Transactions on Dependable and Secure Computing 9 (4) (2011) 494–511.

[33] S. Fortunati, F. Gini, M. S. Greco, A. Farina, A. Graziano, S. Giompapa, An improvement of the state-of-the-art covariance-based methods for statistical anomaly detection algorithms, Signal, Image and Video Processing 10 (4) (2016) 687–694, publisher: Springer.

[34] N. Hoque, D. K. Bhattacharyya, J. K. Kalita, FFSc: a novel measure for low-rate and high-rate DDoS attack detection using multivariate data

analysis, Security and Communication Networks 9 (13) (2016) 2032–2041, publisher: Wiley Online Library.

[35] N. Hoque, H. Kashyap, D. K. Bhattacharyya, Real-time DDoS attack detection using FPGA, Computer Communications 110 (2017) 48–58, publisher: Elsevier.

[36] T. Peng, C. Leckie, K. Ramamohanarao, Proactively detecting distributed denial of service attacks using source IP address monitoring, in: International Conference on Research in Networking, Springer, 2004, pp. 771–782.

[37] J. Udhayan, T. Hamsapriya, Statistical Segregation Method to Minimize the False Detections During DDoS Attacks., IJ Network Security 13 (3) (2011) 152–160.

[38] H. Wang, D. Zhang, K. G. Shin, Detecting SYN flooding attacks, in: Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies, Vol. 3, 2002, pp. 1530–1539.

[39] M. H. Bhuyan, D. K. Bhattacharyya, J. K. Kalita, An empirical evaluation of information metrics for low-rate and high-rate DDoS attack detection, Pattern Recognition Letters 51 (2015) 1–7, publisher: Elsevier.

[40] I. Ozçelik, R. R. Brooks, Cusum-entropy: an efficient method for DDoS attack detection, in: 4th International Istanbul Smart Grid Congress and Fair (ICSG), IEEE, 2016, pp. 1–5.

[41] J. François, I. Aib, R. Boutaba, FireCol: a collaborative protection network for the detection of flooding DDoS attacks, IEEE/ACM Transactions on networking 20 (6) (2012) 1828–1841, publisher: IEEE.

[42] A. Winkelbauer, Moments and Absolute Moments of the Normal Distribution, arXiv:1209.4340v2, 2014.

[43] W. Feller, An Introduction to Probability Theory and Its Application. Volume 2., John Wiley & Sons, 1970.

Appendix A. Proof of Lemma 4.1

Throughout the proof $(U_1, \ldots, U_n)$ are mutually independent rvs with pdf $f_{0,r}$ for U_r. Introduce

$$\rho_r(x) = \frac{f_{1,r}(x)}{f_{0,r}(x)} - 1, \tag{A.1}$$

where $f_{0,r}$ and $f_{1,r}$ are defined in (5) and (8), respectively. The (possibly infinite) constant

$$C_r = \mathbb{E}[\rho_r^2(U_r)]$$

will play a key role in the following. It is known as the Fisher information constant at origin [16] – hereafter simply called the Fisher constant – and can be rewritten as

$$C_r = \int_{\mathbb{R}} \frac{(f_{0,r}(x) - f_{1,r}(x))^2}{f_{0,r}(x)} dx = -1 + \int_{\mathbb{R}} \frac{f_{1,r}^2(x)}{f_{0,r}(x)} dx. \tag{A.2}$$

For later use, notice from (A.1) that

$$\mathbb{E}[\rho_r(Z_r)] = 0. \tag{A.3}$$

Proof of Lemma 4.1. Lemma 4.1 is true if $C_r = \infty$ for some $r \geq 1$. Assume from now on that $\sup_{r\geq 1} C_r < \infty$. We have (cf. (13), (9), (A.1)),

$$\begin{aligned} 2T_V\left(f_0^{(n)}, g^{(n)}\right) &= \int_{\mathbb{R}^n} \Big| \prod_{r=1}^n f_{0,r}(x_r) - \prod_{r=1}^n g_r(x_r) \Big| dx_1 \cdots dx_n \\ &= \int_{\mathbb{R}^n} \Big| 1 - \prod_{r=1}^n (1 + q(n)\rho_r(x_r)) \Big| \prod_{r=1}^n f_0(x_r)\, dx_1 \cdots dx_n \\ &= \mathbb{E}\left[\Big| 1 - \prod_{r=1}^n (1 + q(n)\rho_r(U_r)) \Big| \right]. \end{aligned}$$

Using the inequality $\mathbb{E}[|U|] \leq \sqrt{E[U^2]}$, we obtain

$$\begin{aligned}
\left(2T_V\left(f_0^{(n)}, g^{(n)}\right)\right)^2 &\leq \mathbb{E}\left[\left(1-\prod_{r=1}^{n}(1+q\rho(U_r))\right)^2\right] \\
&= 1-2\mathbb{E}\left[\prod_{r=1}^{n}(1+q\rho_r(U_r)\right] + \mathbb{E}\left[\prod_{r=1}^{n}(1+q(n)\rho(U_r))^2\right] \\
&= 1-2\prod_{r=1}^{n}\mathbb{E}[1+q(n)\rho(U_r)] + \prod_{i=1}^{n}\mathbb{E}\left[(1+q(n)\rho_r(U_r))^2\right] \\
&= \prod_{r=1}^{n}\mathbb{E}\left[(1+q(n)\rho_r(U_r))^2\right] - 1 \quad \text{by using } (A.3) \\
&= \prod_{r=1}^{n}\left(1+q^2(n)\mathbb{E}[\rho_r^2(U_r)]\right) - 1 \quad \text{again by using } (A.3) \\
&= \prod_{r=1}^{n}(1+q^2(n)C_r) - 1,
\end{aligned}$$

which completes the proof. ∎

Appendix B. Proof of Theorem 4.1

In the setting of Theorem 4.1 $f_{0,r}(x) = \frac{e^{-(x-\mu_{0,r})^2/2\sigma_{0,r}^2}}{\sqrt{2\pi\sigma_{0,r}^2}}$ and (obtained from (8) when $I_r = J = 1$, $\mu_{0,1,r} = \mu_{0,r}$, $\sigma_{0,1,r}^2 = \sigma_{0,r}^2$, $\mu_{1,1}(n) = \mu_1(n)$, and $\sigma_{1,1}^2(n) = \sigma_1^2(n)$)

$$f_{1,r}(x,n) = \frac{e^{-\frac{(x-\mu_{0,r}-\mu_1(n))^2}{2(\sigma_{0,r}^2+\sigma_1^2(n))}}}{\sqrt{2\pi(\sigma_{0,r}^2+\sigma_1^2(n))}}.$$

The proof uses Corollary 4.1 in Section 4. We first calculate $C_r(n)$ defined in (20). We obtain

$$\begin{aligned}
C_r(n) &= -1 + \frac{\sqrt{2\pi\sigma_{0,r}^2}}{2\pi(\sigma_{0,r}^2+\sigma_1^2(n))}\int_{\mathbb{R}} e^{-\frac{(x-\mu_{0,r}-\mu_1(n))^2}{\sigma_{0,r}^2+\sigma_1^2(n)}+\frac{(x-\mu_{0,r})^2}{2\sigma_{0,r}^2}}\,dx \\
&= -1 + \frac{\sqrt{2\pi\sigma_{0,r}^2}}{2\pi(\sigma_{0,r}^2+\sigma_1^2(n))} e^{\frac{\mu_1^2(n)}{\sigma_{0,r}^2-\sigma_1^2(n)}} \\
&\quad \times \int_{\mathbb{R}} e^{-\frac{\sigma_{0,r}^2-\sigma_1^2(n)}{2\sigma_{0,r}^2(\sigma_{0,r}^2+\sigma_1^2(n))}\left(x-\frac{\mu_{0,r}(\sigma_{0,r}^2-\sigma_1^2(n))+2\mu_1(n)\sigma_{0,r}^2}{\sigma_{0,r}^2-\sigma_1^2(n)}\right)^2} dx, \qquad \text{(B.1)}
\end{aligned}$$

which is well defined under the second condition in (14). With $\int_{\mathbb{R}} e^{-a(t-b)^2}dt = \sqrt{\pi/a}$ we easily get from (B.1) that

$$C_r(n) \quad = \quad -1 + \frac{\sigma_{0,r}^2}{\sqrt{\sigma_{0,r}^4 - \sigma_1^4(n)}} e^{\frac{\mu_1^2(n)}{\sigma_{0,r}^2 - \sigma_1^2(n)}}. \tag{B.2}$$

Define $h(s,t) = \frac{1}{\sqrt{1-t^2}} e^{\frac{s}{1-t}}$ and note that

$$C_r(n) = -1 + h\left(\mu_1^2(n)/\sigma_{0,r}^2, \sigma_1^2(n)/\sigma_{0,r}^2\right). \tag{B.3}$$

Since h has partial derivatives of all orders in $\mathbb{R} \times (-1,1)$, we know by Taylor theorem that there exists $\theta \in (0,1)$, depending on s and t, such that

$$\begin{aligned}
&h(s,t) = h(0,0) + \frac{\partial}{\partial s}h(0,0)s + \frac{\partial}{\partial t}h(0,0)t \\
&+ \frac{1}{2}\left(\frac{\partial^2}{\partial s^2}h(\theta s, \theta t)s^2 + 2\frac{\partial^2}{\partial s \partial t}h(\theta s, \theta t)st + \frac{\partial^2}{\partial t^2}h(\theta s, \theta t)t^2\right) \\
&= 1 + s + \frac{e^{\frac{\theta s}{1-\theta t}}}{2\sqrt{1-\theta^2 t^2}}\left[\frac{s^2}{(1-\theta t)^2} + \frac{2st}{1-\theta t}\left(\frac{1}{1-\theta^2 t^2} + \frac{1}{1-\theta t} + \frac{\theta s}{(1-\theta t)^2}\right)\right. \\
&\left. + \frac{3\theta^2 t^4}{(1-\theta^2 t^2)^2} - \frac{2\theta^2 s t^3}{(1-\theta^2 t^2)(1-\theta t)^2} - \frac{t^2}{1-\theta^2 t^2} + \frac{2\theta s t^2}{(1-\theta t)^3} + \frac{\theta^2 s^2 t^2}{(1-\theta t)^4}\right] \\
&= 1 + s - \frac{t^2}{2(1-\theta^2 t^2)^{3/2}} e^{\frac{\theta s}{1-\theta t}} + R(s,t),
\end{aligned} \tag{B.4}$$

where

$$\begin{aligned}
R(s,t) := \frac{e^{\frac{\theta s}{1-\theta t}}}{2\sqrt{1-\theta^2 t^2}}&\left[\frac{s^2}{(1-\theta t)^2} + \frac{2st}{1-\theta t}\left(\frac{1}{1-\theta^2 t^2} + \frac{1}{1-\theta t} + \frac{\theta s}{(1-\theta t)^2}\right)\right. \\
&\left. + \frac{3\theta^2 t^4}{(1-\theta^2 t^2)^2} - \frac{2\theta^2 s t^3}{(1-\theta^2 t^2)(1-\theta t)^2} + \frac{2\theta s t^2}{(1-\theta t)^3} + \frac{\theta^2 s^2 t^2}{(1-\theta t)^4}\right].
\end{aligned} \tag{B.5}$$

Define $\sigma_{0,\inf}^2 = \inf_{r \geq 1} \sigma_{0,r}^2$ and $\sigma_{1,\sup}^2 = \sup_{n \geq 1} \sigma_1^2(n)$. Note that $\sigma_{1,\sup}^2 < \sigma_{0,\inf}^2$ under the second condition in (14). By setting $s = \mu_1^2(n)/\sigma_{0,r}^2$ and

$t = \sigma_1^2(n)/\sigma_{0,r}^2$ in (B.4) we obtain from (B.3)

$$
\begin{aligned}
& nq^2(n)C_r(n) = \\
& \frac{nq^2(n)\mu_1^2(n)}{\sigma_{0,r}^2} - nq^2(n)\sigma_1^4(n)\frac{\sigma_{0,r}^2\, e^{\frac{\theta\mu_1^2(n)}{\sigma_{0,r}^2-\theta\sigma_1^2(n)}}}{2(\sigma_{0,r}^4-\theta\sigma_1^4(n))^{\frac{3}{2}}} + nq^2(n)R(\mu_1(n),\sigma_1(n)) \\
& \leq \frac{nq^2(n)\mu_1^2(n)}{\sigma_{0,\inf}^2} + nq^2(n)\sigma_1^4(n)\frac{\sigma_{0,\sup}^2\, e^{\frac{\mu_1^2(n)}{\sigma_{0,\inf}^2-\sigma_{1,\sup}^2}}}{2(\sigma_{0,\inf}^4-\sigma_{1,\sup}^4)^{\frac{3}{2}}} + nq^2(n)|R(\mu_1(n),\sigma_1(n))|.
\end{aligned}
\tag{B.6}
$$

Assumptions in (14)-(15) imply that the first two terms in (B.6) are $\mathcal{O}(1)$; in particular, the first condition in (14) ensures that the exponent of the exponential is $\mathcal{O}(1)$. It is also easily seen that under (14)-(15) $nq^2(n)|R(\mu_1(n),\sigma_1(n))| = \mathcal{O}(1)$, which shows that $nq^2(n)\sup_{1\leq r\leq n} C_r(n) = \mathcal{O}(1)$. The latter necessary implies that $q^2(n)\sup_{1\leq r\leq n} C_r(n) = o(1)$. Hence,

$$
\begin{aligned}
& \sum_{r=1}^{n}\log\left(1+q^2(n)C_r(n)\right) \leq \sum_{r=1}^{n}\log\left(1+q^2(n)\sup_{1\leq r\leq n}C_r(n)\right) \\
& = n\log\left(1+q^2(n)\sup_{1\leq r\leq n}C_r(n)\right) \\
& \sim_n nq^2(n)\sup_{1\leq r\leq n}C_r(n) \quad \text{since } q^2(n)\sup_{1\leq r\leq n}C_r(n) = o(1), \\
& = \mathcal{O}(1)
\end{aligned}
$$

since $nq^2(n)\sup_{1\leq r\leq n} C_r(n) = \mathcal{O}(1)$. The proof is concluded by invoking Corollary 4.1. ∎

Appendix C. Proof of Theorem 4.2

The proof is a generalization of the proof of Theorem 4.1 given in Section Appendix B. It consists of finding the limiting behavior of $nq^2(n)\sup_{1\leq r\leq n} C_r(n)$ as n grows and of applying Corollary 4.1.

With (5)-(8) the Fisher constant C_r in (A.2) writes

$$
C_r(n) = -1 + \int_{\mathbb{R}} \frac{1}{f_{0,r}(x)}\left(\sum_{i=1}^{I_r}\sum_{j=1}^{J}\frac{w_{0,i,r}w_{1,j}e^{-\frac{(x-\mu_{0,i,r}-\mu_{1,j}(n))^2}{2(\sigma_{0,i,r}^2+\sigma_{1,j}^2(n))}}}{\sqrt{2\pi(\sigma_{0,i,r}^2+\sigma_{1,j}^2(n))}}\right)^2 dx. \tag{C.1}
$$

To simplify the notation, from now on we drop the argument n in $\mu_{1,j}(n)$ and $\sigma_{1,j}(n)$. Define

$$h_{x,i,r}(s,t) = \frac{1}{\sqrt{2\pi(\sigma_{0,i,r}^2 + t)}} e^{-\frac{(x-\mu_{0,i,r}-s)^2}{2(\sigma_{0,i,r}^2+t)}}. \tag{C.2}$$

From Taylor theorem there exists $\theta_{x,i,r} \in (0,1)$, hereafter simply denoted by θ, such that

$$\begin{aligned} h_{x,i,r}(s,t) &= h_{x,i,r}(0,0) + \frac{\partial}{\partial s} h_{x,i,r}(0,0)s + \frac{\partial}{\partial t} h_{x,i,r}(0,0)t \\ &+ \frac{1}{2}\frac{\partial^2}{\partial s^2} h_{x,i,r}(\theta s, \theta t)s^2 + \frac{\partial^2}{\partial s \partial t} h_{x,i,r}(\theta s, \theta t)st + \frac{1}{2}\frac{\partial^2}{\partial t^2} h_{x,i,r}(\theta s, \theta t)t^2. \end{aligned} \tag{C.3}$$

This formula holds for all $x, s \in \mathbb{R}$ and for all $t \in \mathbb{R}$ such that $t + \sigma_{0,i,r}^2 > 0$ for $i = 1, \ldots, I_r$, $r \geq 1$, namely, for all $t > -\inf_{1 \leq i \leq I_r, r \geq 1}\{\sigma_{0,i,r}^2\}$. Easy algebra gives

$$h_{x,i,r}(0,0) = \frac{e^{-\frac{(x-\mu_{0,i,r})^2}{2\sigma_{0,i,r}^2}}}{\sqrt{2\pi\sigma_{0,i,r}^2}}, \tag{C.4}$$

$$\frac{\partial}{\partial s} h_{x,i,r}(0,0) = \frac{e^{-\frac{(x-\mu_{0,i,r})^2}{2\sigma_{0,i,r}^2}}}{\sqrt{2\pi}\sigma_{0,i,r}^4}(x - \mu_{0,i,r}),$$

$$\frac{\partial}{\partial t} h_{x,i,r}(0,0) = \frac{e^{-\frac{(x-\mu_{0,i,r})^2}{2\sigma_{0,i,r}^2}}}{2\sqrt{2\pi}\sigma_{0,i,r}^4}\left(\frac{(x-\mu_{0,i,r})^2}{\sigma_{0,i,r}^2} - 1\right), \tag{C.5}$$

$$\frac{\partial^2}{\partial s^2} h_{x,i,r}(\theta s, \theta t) = \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{(x-\mu_{0,i,r}-\theta s)^2}{2(\sigma_{0,i,r}^2+\theta t)}}}{(\sigma_{0,i,r}^2 + \theta t)^{\frac{3}{2}}}\left(\frac{(x-\mu_{0,i,r}-\theta s)^2}{\sigma_{0,i,r}^2 + \theta t} - 1\right), \tag{C.6}$$

$$\begin{aligned} &\frac{\partial^2}{\partial s \partial t} h_{x,i,r}(\theta s, \theta t) = \\ &\frac{1}{2\sqrt{2\pi}} \frac{e^{-\frac{(x-\mu_{0,i,r}-\theta s)^2}{2(\sigma_{0,i,r}^2+\theta t)}}}{(\sigma_{0,i,r}^2 + \theta t)^{\frac{5}{2}}}(x - \mu_{0,i,r} - \theta s)\left(\frac{(x-\mu_{0,i,r}-\theta s)^2}{\sigma_{0,i,r}^2 + \theta t} - 3\right), \end{aligned} \tag{C.7}$$

$$\begin{aligned} &\frac{\partial^2}{\partial t^2} h_{x,i,r}(\theta s, \theta t) = \\ &\frac{1}{2\sqrt{2\pi}} \frac{e^{-\frac{(x-\mu_{0,i,r}-\theta s)^2}{2(\sigma_{0,i,r}^2+\theta t)}}}{(\sigma_{0,i,r}^2 + \theta t)^{\frac{5}{2}}}\left(\frac{(x-\mu_{0,i,r}-\theta s)^4}{2(\sigma_{0,i,r}^2 + \theta t)^2} - \frac{3(x-\mu_{0,i,r}-\theta s)^2}{\sigma_{0,i,r}^2 + \theta t} + \frac{3}{2}\right). \end{aligned} \tag{C.8}$$

Introduce

$$\Delta_r(x,n) := \sum_{i=1}^{I_r}\sum_{j=1}^{J} w_{0,i,r}w_{1,j}\left(\frac{1}{2}\frac{\partial^2}{\partial s^2}h_{x,i,r}(\theta\mu_{1,j},\theta\sigma^2_{1,j})\mu^2_{1,j}\right.$$

$$\left. + \frac{\partial^2}{\partial s\partial t}h_{x,i,r}(\theta\mu_{1,j},\theta\sigma^2_{1,j})\mu_{1,j}\sigma^2_{1,j} + \frac{1}{2}\frac{\partial^2}{\partial t^2}h_{x,i,r}(\theta\mu_{1,j},\theta\sigma^2_{1,j})\sigma^4_{j,1}\right), \tag{C.9}$$

$$\alpha_r(x) := \sum_{i=1}^{I_r}\frac{w_{0,i,r}}{\sigma^4_{i,0}\sqrt{2\pi}}\,(x-\mu_{0,i,r})\,e^{-\frac{(x-\mu_{0,i,r})^2}{2\sigma^2_{0,i,r}}}, \tag{C.10}$$

$$\beta_r(x) := \frac{1}{2}\sum_{i=1}^{I_r}\frac{w_{0,i,r}}{\sigma^4_{i,0}\sqrt{2\pi}}\left(\frac{(x-\mu_{0,i,r})^2}{\sigma^2_{0,i,r}}-1\right)e^{-\frac{(x-\mu_{0,i,r})^2}{2\sigma^2_{0,i,r}}}. \tag{C.11}$$

We have

$$\begin{aligned}
&\sum_{i=1}^{I_r}\sum_{j=1}^{J}\frac{w_{0,i,r}w_{1,j}}{\sqrt{2\pi(\sigma^2_{0,i,r}+\sigma^2_{1,j})}}e^{-\frac{(x-\mu_{0,i,r}-\mu_{1,j})^2}{2(\sigma^2_{0,i,r}+\sigma^2_{1,j})}}\\
&= \sum_{i=1}^{I_r}\sum_{j=1}^{J} w_{0,i,r}w_{1,j}h_{x,i,r}(\mu_{1,j},\sigma^2_{1,j}) \quad \text{by } (C.2),\\
&= f_{0,r}(x)+\alpha_r(x)\mu_1(n)+\beta_r(x)\sigma^2_1(n)+\Delta_r(x,n),
\end{aligned} \tag{C.12}$$

by using (C.4)-(C.11), so that by (C.1)

$$\begin{aligned}
C_r(n) &= -1+\int_{\mathbb{R}}\frac{1}{f_{0,r}(x)}(f_{0,r}(x)+\alpha_r(x)\mu_1(n)+\beta_r(x)\sigma^2_1(n)+\Delta_r(x,n))^2dx\\
&= \mu^2_1(n)\int_{\mathbb{R}}\frac{\alpha^2_r(x)}{f_{0,r}(x)}dx+\sigma^4_1(n)\int_{\mathbb{R}}\frac{\beta^2_r(x)}{f_{0,r}(x)}dx\\
&+ 2\mu_1(n)\sigma^2_1(n)\int_{\mathbb{R}}\frac{\alpha_r(x)\beta_r(x)}{f_{0,r}(x)}dx+\sigma^4_1(n)K_r(n),
\end{aligned} \tag{C.13}$$

where

$$K_r(n) := \frac{1}{\sigma^4_1(n)}\int_{\mathbb{R}}\frac{\Delta_r(x,n)}{f_{0,r}(x)}(2f_{0,r}(x)+2\mu_1(n)\alpha_r(x)+2\sigma^2_1(n)\beta_r(x)+\Delta_r(x,n))dx. \tag{C.14}$$

To derive (C.13) we have used that $\int_{\mathbb{R}}\alpha_r(x)dx = \int_{\mathbb{R}}\beta_r(x)dx = 0$. Therefore,

$$\sup_{1\le r\le n} C_r(n) \le \mu^2_1(n)\sup_{1\le r\le n}\int_{\mathbb{R}}\frac{\alpha^2_r(x)}{f_{0,r}(x)}dx+\sigma^4_1(n)\sup_{1\le r\le n}\int_{\mathbb{R}}\frac{\beta^2_r(x)}{f_{0,r}(x)}dx \tag{C.15}$$

$$+ 2\mu_1(n)\sigma^2_1(n)\sup_{1\le r\le n}\left|\int_{\mathbb{R}}\frac{\alpha_r(x)\beta_r(x)}{f_{0,r}(x)}dx\right|+\sigma^4_1(n)\sup_{1\le r\le n}|K_r(n)|. \tag{C.16}$$

It is shown in Appendix Appendix D that the coefficients of $\mu_1^2(n)$ and $\sigma_1^4(n)$ in (C.15) and the coefficient of $\mu_1(n)\sigma_1^2(n)$ in (C.16) are all $\mathcal{O}(1)$ quantities, and in Appendix Appendix E that $\sup_{1\leq r\leq n}|K_r(n)| = \mathcal{O}(1)$. Hence, by multiplying both sides of (C.15)-(C.16) by $nq^2(n)$ and by (17) we obtain that $nq^2(n)\sup_{1\leq r\leq n} C_r(n) = \mathcal{O}(1)$. The same argument used to conclude the proof of Theorem 4.1 in Section Appendix B can be duplicated to conclude the proof of Theorem 4.2.

Appendix D. Uniform boundedness of the three integrals in (C.13)

Define $a_{i,r} = \frac{w_{0,i,r}}{\sqrt{2\pi\sigma^2_{0,i,r}}}$, $b_{i,r} = \frac{w_{0,i,r}}{\sigma^4_{0,i,r}\sqrt{2\pi}}$, and let $i_r^\star := \text{argmax}_{i=1,\ldots,I_r}\{\sigma^2_{0,i,r}\}$.

Since $\frac{w_{0,i,r}}{\sqrt{2\pi\sigma^2_{0,i,r}}} e^{-\frac{(x-\mu_{i,0,r})^2}{2\sigma^2_{0,i,r}}} \geq 0$ for all $x\in\mathbb{R}$ and $i = 1,\ldots,I_r$, the definition of $f_{0,r}(x)$ in (5) implies

$$f_{0,r}(x) \geq a_{i_r^\star,r} e^{-\frac{(x-\mu_{0,i_r^\star,r})^2}{2\sigma^2_{0,i_r^\star,r}}}, \quad \forall x\in\mathbb{R}. \tag{D.1}$$

Hence,

$$\begin{aligned}
&\int_{\mathbb{R}} \frac{\alpha_r(x)^2}{f_{0,r}(x)}dx \\
&= \int_{\mathbb{R}} \frac{1}{f_{0,r}(x)} \sum_{1\leq i,l\leq I_r} b_{i,r}b_{l,r}(x-\mu_{0,i,r})(x-\mu_{0,l,r}) e^{-\frac{(x-\mu_{0,i,r})^2}{2\sigma^2_{0,i,r}} - \frac{(x-\mu_{0,l,r})^2}{2\sigma^2_{0,l,r}}} dx \\
&\leq \frac{1}{a_{i_r^\star,r}} \sum_{1\leq i,l\leq I_r} b_{i,r}b_{l,r} \int_{\mathbb{R}} (x-\mu_{0,i,r})(x-\mu_{0,l,r}) \\
&\times e^{\frac{-(x-\mu_{0,i,r})^2}{2\sigma^2_{0,i,r}} - \frac{(x-\mu_{0,l,r})^2}{2\sigma^2_{0,l,r}} + \frac{(x-\mu_{0,i_r^\star,r})^2}{2\sigma^2_{0,i_r^\star,r}}} dx.
\end{aligned} \tag{D.2}$$

The coefficient of x^2 in the exponential in (D.2) is $-\frac{1}{2}(\sigma^2_{0,i_r^\star,r}(\sigma^2_{0,i,r}+\sigma^2_{0,l,r}) - \sigma^2_{0,i,r}\sigma^2_{0,l,r})$; it is strictly negative from the definition of $i_r^\star$ since $\sigma^2_{0,i_r^\star,r} \geq \sigma^2_{0,i,r} \geq \sigma^2_{0,i,r} \times \frac{\sigma^2_{0,l,r}}{\sigma^2_{0,i,r}+\sigma^2_{0,l,r}}$ for all i,l. This shows that the integral $\int_{\mathbb{R}} \frac{\alpha_r(x)^2}{f_{0,r}(x)}dx$ is finite for each $r\geq 1$. Under conditions in (10) and their consequences in (12) it is easily seen that $\sup_{1\leq r\leq n, n\geq 1}\int_{\mathbb{R}} \frac{\alpha_r(x)^2}{f_{0,r}(x)}dx = \mathcal{O}(1)$.

Similarly, one can show that the second and third integrals in the r.h.s. of (C.13) are uniformly bounded for $n\geq 1$.

Appendix E. Proof that $\sqrt{\sup_{1\leq r\leq n}|K_r(n)|}=\mathcal{O}(1)$

$\Delta(x,n)$ defined in (C.9) can be written as

$$\Delta_r(x,n)=\Delta_{1,r}(x,n)+\Delta_{2,r}(x,n)+\Delta_{3,r}(x,n), \tag{E.1}$$

with

$$\begin{aligned}
\Delta_{1,r}(x,n) &:= \frac{1}{2\sqrt{2\pi}}\sum_{i=1}^{I_r}w_{0,i,r}\sum_{j=1}^{J}w_{1,j}\mu_{1,j}^2\frac{e^{-\frac{(x-\mu_{0,i,r}-\theta\mu_{1,j})^2}{2(\sigma_{0,i,r}^2+\theta\sigma_{1,j}^2)}}}{(\sigma_{0,i,r}^2+\theta\sigma_{1,j}^2)^{\frac{3}{2}}}\\
&\quad\times\left(\frac{(x-\mu_{0,i,r}-\theta\mu_{1,j})^2}{\sigma_{0,i,r}^2+\theta\sigma_{1,j}^2}-1\right)\\
\Delta_{2,r}(x,n) &:= \frac{1}{2\sqrt{2\pi}}\sum_{i=1}^{I_r}w_{0,i,r}\sum_{j=1}^{J}w_{1,j}\mu_{1,j}\sigma_{1,j}^2\frac{e^{-\frac{(x-\mu_{0,i,r}-\theta\mu_{1,j})^2}{2(\sigma_{0,i,r}^2+\theta\sigma_{1,j}^2)}}}{(\sigma_{0,i,r}^2+\theta\sigma_{1,j}^2)^{\frac{5}{2}}}\\
&\quad\times(x-\mu_{0,i,r}-\theta\mu_{1,j})\left(\frac{(x-\mu_{0,i,r}-\theta\mu_{1,j})^2}{\sigma_{0,i,r}^2+\theta\sigma_{1,j}^2}-3\right)\\
\Delta_{3,r}(x,n) &:= \frac{1}{4\sqrt{2\pi}}\sum_{i=1}^{I_r}w_{0,i}\sum_{j=1}^{J}w_{1,j}\sigma_{1,j}^4\frac{e^{-\frac{(x-\mu_{0,i,r}-\theta\mu_{1,j})^2}{2(\sigma_{0,i,r}^2+\theta\sigma_{1,j}^2)}}}{(\sigma_{0,i,r}^2+\theta\sigma_{1,j}^2)^{\frac{5}{2}}}\\
&\quad\times\left(\frac{(x-\mu_{0,i,r}-\theta\mu_{1,j})^4}{2\sigma_{0,i,r}^4}-\frac{3(x-\mu_{0,i,r}-\theta\mu_{1,j})^2}{\sigma_{0,i,r}^2+\theta\sigma_{1,j}^2}+\frac{3}{2}\right).
\end{aligned}$$

Introducing (E.1) into (C.14) gives

$$\begin{aligned}
&K_r(n)\\
&=\frac{1}{\sigma_1(n)^4}\int_{\mathbb{R}}\frac{\Delta_{1,r}(x,n)}{f_{0,r}(x)}[2f_{0,r}(x)+2\mu_1(n)\alpha(x)+2\sigma_1^2(n)\beta(x)+\Delta_r(x,n)]dx\\
&+\frac{1}{\sigma_1(n)^4}\int_{\mathbb{R}}\frac{\Delta_{2,r}(x,n)}{f_{0,r}(x)}[2f_{0,r}(x)+2\mu_1(n)\alpha(x)+2\sigma_1^2(n)\beta(x)+\Delta_r(x,n)]dx\\
&+\frac{1}{\sigma_1(n)^4}\int_{\mathbb{R}}\frac{\Delta_{3,r}(x,n)}{f_{0,r}(x)}[2f_{0,r}(x)+2\mu_1(n)\alpha(x)+2\sigma_1^2(n)\beta(x)+\Delta_r(x,n)]dx.
\end{aligned} \tag{E.2}$$

Denote by $J_{1,r}(n)$, $J_{2,r}(n)$, and $J_{3,r}(n)$ the three integrals in the r.h.s. of (E.2). By using the bound in (D.1) we obtain

$$|J_{1,r}(n)| \leq \frac{1}{\sigma_1(n)^4} \cdot \frac{\sqrt{2\pi\sigma^2_{0,i^\star_r,r}}}{w_{0,i^\star_r,r}} \cdot \frac{1}{2\sqrt{2\pi}} \sum_{i=1}^{I_r} \frac{w_{0,i,r}}{\sigma^3_{0,i,r}} \sum_{j=1}^{J} w_{1,j}\mu_{1,j}(n)^2$$
$$\times \left[\int_{\mathbb{R}} |2f_{0,r}(x) + 2\mu_1(n)\alpha(x) + 2\sigma_1^2(n)\beta(x) + \Delta_r(x,n)| \right.$$
$$\left. \times\, e^{-\frac{(x-\mu_{0,i,r}-\theta\mu_{1,j}(n))^2}{2(\sigma^2_{0,i,r}+\theta\sigma_1^2(n))} + \frac{(x-\mu_1(n))^2}{2\sigma^2_{0,i^\star_r,r}}} \left(\frac{(x-\mu_{0,i,r}-\theta\mu_1(n))^2}{\sigma^2_{0,i,r}} + 1 \right) dx \right]. \quad \text{(E.3)}$$

Under the second condition in (16), the coefficients of x^2 in the various exponentials in (E.3) are always strictly negative, which shows that the integral in (E.3) is finite for every r and n. We conclude from that and conditions in (10)-(11) and in (16) that $\sup_{1\leq r\leq n} |J_{1,r}(n)| = \mathcal{O}(1)$.

A similar analysis yields $\sup_{1\leq r\leq n} |J_{2,r}(n)| = \mathcal{O}(1)$ and $\sup_{1\leq r\leq n} |J_{3,r}(n)| = \mathcal{O}(1)$.

Appendix F. $\mathbb{E}[|U - \mathbb{E}[U]|^3]$ if U is a mixture of Gaussians

Assume that the pdf of U is

$$u(x) = \sum_{i=1}^{I} \frac{w_j}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu_j)^2}{2\sigma_i^2}},$$

with $0 < w_i < 1$, $\sum_{i=1}^I q_i = 1$, and $\sigma_i > 0$. Note that $\mathbb{E}[U] = \sum_{i=1}^I w_i\mu_i := \mu$.

Let V be a Gaussian rv with mean α and variance β^2. Then [42, Formula (18)]

$$\mathbb{E}[|V-\alpha|^\nu] = \sqrt{\frac{2^\nu\beta^{2\nu}}{\pi}}\Gamma\left(\frac{\nu+1}{2}\right),$$

for $\nu > -1$, where $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$. In particular,

$$\mathbb{E}[|V-\alpha|] = \sqrt{\frac{2\beta^2}{\pi}}, \quad \mathbb{E}[|Z-\alpha|^2] = \beta^2, \quad \mathbb{E}[|Z-\alpha|^3] = 2\sqrt{\frac{2\beta^6}{\pi}}. \quad \text{(F.1)}$$

Therefore (Hint: $|a+b|^3 \leq |a|^3 + 3a^2|b| + 3|a|b^2 + |b|^3$),

$$\begin{aligned}
\mathbb{E}[|U-\mu|^3] &= \sum_{i=1}^{I} \frac{w_i}{\sqrt{2\pi\sigma_i^2}} \int_{-\infty}^{\infty} |x-\mu|^3 e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} dx \\
&= \sum_{i=1}^{I} \frac{w_i}{\sqrt{2\pi\sigma_i^2}} \int_{-\infty}^{\infty} |(x-\mu_i)+(\mu_i-\mu)| e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} dx \\
&\leq \sum_{i=1}^{I} \frac{w_i}{\sqrt{2\pi\sigma_i^2}} \int_{-\infty}^{\infty} |x-\mu_i|^3 e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} dx \\
&+ 3\sum_{i=1}^{I} \frac{w_i}{\sqrt{2\pi\sigma_i^2}} |\mu_i-\mu| \int_{-\infty}^{\infty} (x-\mu_i)^2 e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} dx \\
&+ 3\sum_{i=1}^{I} \frac{w_i}{\sqrt{2\pi\sigma_i^2}} (\mu_i-\mu)^2 \int_{-\infty}^{\infty} |x-\mu_i| e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} dx \\
&+ \sum_{i=1}^{I} \frac{w_i}{\sqrt{2\pi\sigma_i^2}} |\mu_i-\mu|^3 \int_{-\infty}^{\infty} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} dx \\
&= 2\sqrt{\frac{2}{\pi}} \sum_{i=1}^{I} w_i \sqrt{\sigma_i^6} + 3\sum_{i=1}^{I} w_i |\mu_i-\mu| \sigma_i^2 \\
&+ 3\sqrt{\frac{2}{\pi}} \sum_{i=1}^{I} w_i (\mu_i-\mu)^2 \sqrt{\sigma_i^2} + \sum_{i=1}^{I} w_i |\mu_i-\mu|^3 < \infty.
\end{aligned}$$

Appendix G. Proof of Theorem 4.4

The proof uses the Berry-Esseen theorem (see e.g. [43, Theorem 2, p. 544]) which we state below for sake of completeness.

Lemma Appendix G.1 (Berry-Esseen theorem)**.**

Let $W_1, \ldots, W_n$ be mutually independent rvs with finite expectation, and strictly positive and finite variance, and assume that $\mathbb{E}[|W_i - \mathbb{E}[W_i]|^3]$ is finite for $i = 1, \ldots, n$. For all x and n,

$$\left| \mathbb{P}\left(\frac{\sum_{i=1}^{n}(W_i - \mathbb{E}[W_i])}{\sqrt{\sum_{i=1}^{n} var(W_i)}} < x \right) - \Phi(x) \right| \leq \frac{6\sum_{i=1}^{n} \mathbb{E}[|W_i - \mathbb{E}[W_i]|^3]}{\left(\sum_{i=1}^{n} var(W_i)\right)^{3/2}},$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt$ is the cdf of the standard normal distribution.

Denote by $\bar{\Phi}(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{1}{2}t^2} dt$ the ccdf of the standard normal distribution.

Proof of Theorem 4.4: Recall that Z_r is the upload traffic generated by home $r = 1, \ldots, n$ in a given time window. Define $\bar{Z}_n = \frac{Z_1 + \cdots + Z_n}{n}$.

The detector is $\bar{z}_n < M(n) + U$ under H_0 and $\bar{z}_n > M(n) + U$ under H_1, with $\bar{z}_n$ the observed value of the rv $\bar{Z}_n$ and $M(n) := \frac{1}{n}\sum_{r=1}^{n} \mu_{0,r}$ the average regular traffic generated by a home in a given time window. We have

$$\begin{aligned}
p_{FA} &= \mathbb{P}(\bar{Z}_n > M(n) + U \,|\, H_0) \\
&= \mathbb{P}\left(\frac{1}{n}\sum_{r=1}^{n} X_r > M(n) + U \right) \\
&= \mathbb{P}\left(\frac{\sum_{r=1}^{n}(X_r - \mu_{0,r})}{\sqrt{\sum_{r=1}^{n}\sigma_{0,r}^2}} > \frac{nU}{\sqrt{\sum_{r=1}^{n}\sigma_{0,r}^2}} \right) \\
&\leq \left| \mathbb{P}\left(\frac{\sum_{r=1}^{n}(X_r - \mu_{0,r})}{\sqrt{\sum_{r=1}^{n}\sigma_{0,r}^2}} > \frac{nU}{\sqrt{\sum_{r=1}^{n}\sigma_{0,r}^2}} \right) - \bar{\Phi}\left(\frac{nU}{\sqrt{\sum_{r=1}^{n}\sigma_{0,r}^2}} \right) \right| \\
&\quad + \bar{\Phi}\left(\frac{nU}{\sqrt{\sum_{r=1}^{n}\sigma_{0,r}^2}} \right).
\end{aligned}$$

Under Assumptions in (23) and the strict positiveness of $\sigma_{0,r}^2$ the Berry-Esseen Theorem applies to $\{X_r\}_r$, to give

$$p_{FA} \leq \frac{6\sum_{r=1}^{n}\mathbb{E}[|X_r - \mu_{0,r}|^3]}{\left(\sum_{r=1}^{n}\sigma_{0,r}^2\right)^{3/2}} + \bar{\Phi}\left(\frac{nU}{\sqrt{\sum_{r=1}^{n}\sigma_{0,r}^2}} \right). \tag{G.1}$$

Since

$$\frac{\sum_{r=1}^{n}\mathbb{E}[|X_r - \mu_{0,r}|^3]}{\left(\sum_{r=1}^{n}\sigma_{0,r}^2\right)^{3/2}} = \frac{1}{\sqrt{n}} \times \frac{\frac{1}{n}\sum_{r=1}^{n}\mathbb{E}[|X_r - \mu_{0,r}|^3]}{\left(\frac{1}{n}\sum_{r=1}^{n}\sigma_{0,r}^2\right)^{3/2}}, \tag{G.2}$$

the last two conditions in (23) imply that the r.h.s. of (G.2) can be made arbitrarily small by increasing n. Therefore, for any $\alpha \in (0,1)$ there exists n_1 such that

$$p_{FA} < \frac{\alpha}{2} + \bar{\Phi}\left(\frac{nU}{\sqrt{\sum_{r=1}^{n}\sigma_{0,r}^2}} \right) \tag{G.3}$$

for all $n > n_1$. Using now the inequality $\bar{\Phi}(x) \leq \frac{1}{\sqrt{2\pi}}\frac{e^{-\frac{1}{2}x^2}}{x}$, we get from (G.1) that

$$p_{FA} \leq \frac{\alpha}{2} + \frac{1}{\sqrt{2\pi}}\frac{\sqrt{V(n)}}{\sqrt{n}U}e^{-\frac{nU^2}{2V(n)}},$$

with $V(n) := \frac{1}{n}\sum_{r=1}^{n}\sigma_{0,r}^2$.

For n fixed, the equation $\frac{\alpha}{2} = \frac{1}{\sqrt{2\pi}}\frac{\sqrt{V(n)}}{y}e^{-\frac{y^2}{2V(n)}}$ has a single root in $y \in (0,\infty)$. Call it $c(n)$. Existence and uniqueness of the solution follow from the fact that the mapping $y \to \frac{1}{\sqrt{2\pi}}\frac{\sqrt{V(n)}}{y}e^{-\frac{y^2}{2V(n)}}$ is strictly decreasing in $(0,\infty)$ with $\lim_{y\to 0}\frac{1}{\sqrt{2\pi}}\frac{\sqrt{V(n)}}{y}e^{-\frac{y^2}{2V(n)}} = +\infty$ and $\lim_{y\to+\infty}\frac{1}{\sqrt{2\pi}}\frac{\sqrt{V(n)}}{y}e^{-\frac{y^2}{2V(n)}} = 0$.

Take $U = \frac{c(n)}{\sqrt{n}}$. Then,

$$p_{FA} \leq \alpha,$$

for $n > n_1$.

Observe that $\sup_{n\geq 1} c(n) < \infty$. Indeed, by definition of $c(n)$,

$$\frac{\alpha}{2} = \frac{1}{\sqrt{2\pi}}\frac{\sqrt{V(n)}}{c(n)}e^{-\frac{c(n)^2}{2V(n)}}. \tag{G.4}$$

If $\lim c(n) = \infty$ the second condition in (23) would imply that the r.h.s. of (G.4) goes to 0 as $n \to \infty$, which would contradict the fact that $c(n)$ solves (G.4) for all n. This shows that $\sup_{n\geq 1} c(n) < \infty$.

Let $W_r := X_r + \chi_r Y_r$. Recall that the quantities $\mathbb{E}[X_r] = \mu_{0,r}$, $\mathrm{var}(X_r) = \sigma_{0,r}^2$, $\mathbb{E}[|X_r - \mu_{0,r}|^3]$, $\mathbb{E}[Y_r] = \mu_1(n)$, $\mathrm{var}(Y_r)$, and $\mathbb{E}[|Y_r - \mu_1(n)|^3]$ are finite, that the rvs X_r and Y_r are independent, and that conditions (23) and (24) are assumed to hold.

Let us now focus on p_{MD}, the probability of miss-detection. It is given by

$$\begin{aligned} p_{MD} &= \lim_n \mathbb{P}\left(\bar{Z}_n < M(n) + \frac{c(n)}{\sqrt{n}} \mid H_1\right) \\ &= \lim_n \mathbb{P}\left(\frac{\sum_{r=1}^n W_r}{n} < M(n) + \frac{c(n)}{\sqrt{n}}\right) \\ &= \lim_n \mathbb{P}\left(\frac{\sum_{r=1}^n (W_r - \mathbb{E}[W_r]))}{\sqrt{\sum_{r=1}^n \mathrm{var}(W_r)}} < \frac{\sqrt{n}c(n) - nq(n)\mu_1(n)}{\sqrt{\sum_{r=1}^n \mathrm{var}(W_r)}}\right), \end{aligned}$$

by using $\mathbb{E}[W_r] = \mu_{0,r} + q(n)\mu_1(n)$.

Let us show that the Berry-Esseen Theorem applies to the rvs $\{W_r\}_r$. We have $\mathbb{E}[W_r] = \mu_{0,r} + q(n)\mu_1(n)$ which is finite and $\mathrm{var}(W_r) = \sigma_{0,r}^2 + \mathrm{var}(\chi_r Y_r)$

which is also finite since $\mathrm{var}(Y_r)$ is finite and $\chi_r \in \{0,1\}$. Last,

$$\begin{aligned}\mathbb{E}[|W_r - \mathbb{E}[W_r]|^3| &= \mathbb{E}[|(X_r - \mu_{0,r}) + (\chi_r Y_r - q(n)\mu_1(n))|^3 \\ &\leq \mathbb{E}[|X_r - \mu_{0,r}|^3] + 3\mathrm{var}(X_r)\mathbb{E}[|\chi_r Y_r - q(n)\mu_1(n)|] \\ &\quad + 3\mathbb{E}[|X_r - \mu_{0,r}|]\mathrm{var}(\chi_r Y_r) + \mathbb{E}[|\chi_r Y_r - q(n)\mu_1(n)|^3]\end{aligned}$$

by using the inequality $|a+b|^3 \leq |a|^3 + 3a^2|b| + 3|a|b^2 + |b|^3$. From the inequality $\mathbb{E}[|U|] \leq \sqrt{\mathbb{E}[U^2]}$ we find

$$\begin{aligned}\mathbb{E}[|W_r - \mathbb{E}[W_r]|^3] &\leq \mathbb{E}[|X_r - \mu_{0,r}|^3] + 3\mathrm{var}(X_r)\sqrt{\mathrm{var}(\chi_r Y_r)} \\ &\quad + 3\sqrt{\mathrm{var}(X_r)}\mathrm{var}(\chi_r Y_r) + \mathbb{E}[|\chi_r Y_r - q(n)\mu_1(n)|^3] \qquad \text{(G.5)} \\ &= \mathcal{O}(1) + \mathbb{E}[|\chi_r Y_r - q(n)\mu_1(n)|^3] \qquad \text{(G.6)}\end{aligned}$$

by using the second condition in (23) (which says that $\sup_n \mathrm{var}(X_r) < \infty$) and the second condition in (24). Let us show that $\mathbb{E}[|\chi_r Y_r - q(n)\mu_1(n)|^3])$ is finite. We have (Hint: $\mathrm{var}(U+a) = \mathrm{var}(U)$ for any constant a and $\mathbb{E}[|U] \leq \sqrt{\mathbb{E}[U^2]}$)

$$\begin{aligned}&\mathbb{E}[|\chi_r Y_r - q(n)\mu_1(n)|^3] = \mathbb{E}[|(\chi_r(Y_r - \mu_r) + (\chi_r - q(n))\mu_1(n)|^3] \\ &\leq q(n)\mathbb{E}[|Y_r - \mu_1(n)|^3] + 3q(n)\mu_1(n)\mathrm{var}(Y_r)\mathbb{E}[|\chi_r - q(n)|] \\ &+ 3q(n)\mu_1^2(n)\sqrt{\mathrm{var}(Y_r)}\mathbb{E}[(\chi_r - q(n))^2] + q^3(n)\mu_1^3(n)\mathbb{E}[|\chi_r - q(n)|^3] \qquad \text{(G.7)}\end{aligned}$$

which is finite since $|\chi_r - q(n)| \leq 1$. This proves that $\mathbb{E}[|W_r - \mathbb{E}[W_r]|^3]$ is finite and shows that the Berry-Esseen Theorem applies to the rvs $\{W_r\}_r$.

Similarly to the derivation of (G.1), Berry-Esseen inequality yields

$$\begin{aligned}p_{MD} &\leq \lim_n \frac{6\sum_{r=1}^n \mathbb{E}[|W_r - \mathbb{E}[W_r]|^3]}{\left(\sum_{r=1}^n \mathrm{var}(W_r)\right)^{3/2}} + \lim_n \phi\left(\frac{\sqrt{n}c(n) - nq(n)\mu_1(n)}{\sqrt{\sum_{r=1}^n \mathrm{var}(W_r)}}\right) \\ &= \lim_n \frac{6}{\sqrt{n}} \times \frac{\frac{1}{n}\sum_{r=1}^n \mathbb{E}[|W_r - \mathbb{E}[W_r]|^3]}{\left(\frac{1}{n}\sum_{r=1}^n \sigma_{0,r}^2 + \mathrm{var}(\chi_r Y_r)\right)^{3/2}} \\ &\quad + \lim_n \phi\left(\frac{c(n) - \sqrt{n}q(n)\mu_1(n)}{\sqrt{\frac{1}{n}\sum_{r=1}^n \sigma_{0,r}^2 + \mathrm{var}(\chi_r Y_r)}}\right). \qquad \text{(G.8)}\end{aligned}$$

The second condition in (23) together with the finiteness of $\sup_{n\geq 1} c(n)$ shown above, shows that

$$\lim_n \left(\frac{c(n) - \sqrt{n}q(n)\mu_1(n)}{\sqrt{\frac{1}{n}\sum_{r=1}^n \sigma_{0,r}^2 + q(n)\sigma_1^2(n) + q(n)(1-q(n))\mu_1^2(n)}}\right) = -\infty$$

when both conditions in (24) hold; hence

$$\lim_n \phi\left(\frac{c(n)-\sqrt{n}q(n)\mu_1(n)}{\sqrt{\frac{1}{n}\sum_{r=1}^n \sigma_{0,r}^2 + q(n)\sigma_1^2(n) + q(n)(1-q(n))\mu_1^2(n)}}\right) = \phi(-\infty) = 0, \tag{G.9}$$

by continuity of the mapping ϕ.

We now show that the term $\frac{\frac{1}{n}\sum_{r=1}^n \mathbb{E}[|W_r - \mathbb{E}[W_r]|^3]}{\left(\frac{1}{n}\sum_{r=1}^n \sigma_{0,r}^2 + \mathrm{var}(\chi_r Y_r)\right)^{3/2}}$ in (G.8) is finite. The second conditions in (23) and (24) imply that the denominator is $\mathcal{O}(1)$. By (G.5),

$$\begin{aligned}
&\frac{1}{n}\sum_{r=1}^n \mathbb{E}[|W_r - \mathbb{E}[W_r]|^3] \leq \frac{1}{n}\sum_{r=1}^n \mathbb{E}[|X_r - \mu_{0,r}|^3] + \frac{3}{n}\sum_{r=1}^n \mathrm{var}(X_r) \times \sqrt{\mathrm{var}(\chi_r Y_r)} \\
&+ \frac{3}{n}\sum_{r=1}^n \sqrt{\mathrm{var}(X_r)} \times \mathrm{var}(\chi_r Y_r) + \mathbb{E}[|\chi_r Y_r - q(n)\mu_1(n)|^3] \\
&= \mathcal{O}(1) \times \left(1 + \frac{3}{n}\sum_{r=1}^n \sigma_{0,r}\right)
\end{aligned} \tag{G.10}$$

from the second and third conditions in (23) and the second condition in (24), and where the finiteness of $\mathbb{E}[|\chi_r Y_r - q(n)\mu_1(n)|^3]$ was shown in (G.7). From the inequality $\sqrt{x} \leq 1 + x$, we conclude from the second condition in (23) and (G.10) that $\frac{1}{n}\sum_{r=1}^n \mathbb{E}[|W_r - \mathbb{E}[W_r]|^3] = \mathcal{O}(1)$. This concludes the proof that $\lim_n p_{MD} = 0$ and proves the theorem. ∎