

Detection of Interest Flooding Attacks in Named Data Networking using Hypothesis Testing

Tan N. Nguyen, Rémi Cogranne, Guillaume Doyen and Florent Retraint

ICD - STMR - UMR 6281 CNRS

Troyes University of Technology

Troyes, France

ngoc_tan.nguyen@utt.fr ; remi.cogranne@utt.fr ; guillaume.doyen@utt.fr ; florent.retraint@utt.fr

Abstract—With the rapid growth of Internet traffic, new emerging network architectures are under deployment. Those architectures will substitute the current IP/TCP network only if they can ensure better security. Currently, the most advanced proposal for future Internet architecture is Named Data Networking (NDN). However, new computer network architectures bring new types of attacks. This paper focuses on the detection against Interest flooding - one of the most threatening attacks in NDN. The statistical detection is studied within the framework of hypothesis testing. First, we address the case in which all traffic parameters are known. In this context, the optimal test is designed and its statistical performance is given. This allows providing an upper bound on the highest detection accuracy one can expect. Then, a linear parametric model is proposed to estimate unknown parameters and to design a practical test for which the statistical performance is also provided. Numerical results show the relevance of the proposed methodology.

Index Terms—Network security, Named Data Networking, Interest flooding, Statistical detection, Hypothesis testing.

I. INTRODUCTION

Internet usage still keeps growing tremendously, challenging the current IP network with many emerging usages for which it has not been designed, e.g. handling huge content distribution access from users and maintaining connection for mobile devices. Therefore, a current important research topic focuses on proposing clean-slate network architectures which face the future Internet requirements. Such new network designs have been proposed in *Information Centric Network* (ICN) [1] architectures. Among ICN proposals, *Named Data Networking* (NDN) [2] is considered to be the most promising future network. This architecture draws a lot of attention from the research community. NDN testbeds have been deployed and shared between institutions from America, Europe and Asia. Moreover, many telco operators are also interested in this proposal, with many testbed deployment projects (e.g. DOCTOR project) to investigate its feasibility.

Though NDN architecture is currently rather complete, its implementation is still under development. Research efforts on NDN are focusing on management, monitoring and, especially, security. Each component in NDN architecture possibly becomes the target of new attacks. In this paper, we focus on the *Interest flooding attack* (IFA) [3]. This attack can be launched easily, without much requirements of knowledge, while potentially causing large scale damage on network availability.

This paper studies the statistical detection of IFA. The problem is cast within the framework of hypothesis testing theory, which to the best of our knowledge, has never been studied in this context. The main contributions of this paper are briefly summarized as follows. First, the optimal *Likelihood Ratio Test* (LRT) is designed in the theoretical case of a perfectly known legitimate traffic. The optimality of this statistical test is ensured whatever the attack payload may be. This test serves as an upper bound of the detection accuracy one can expect for IFA detection. Secondly, in a scenario where the legitimate traffic is unknown, we propose a parametric statistical model upon which a practical *Generalized LRT* (GLRT) is designed. Finally, the statistical properties of the proposed GLRT are established analytically. This especially allows guaranteeing a prescribed false-alarm probability and computing the detection accuracy.

The paper is organized as follows. Section II recalls the NDN architecture and provides an overview of IFA in NDN. Next, Section III formalizes the problem of IFA detection within hypothesis testing theory. The optimal LRT and its statistical properties are presented in Section IV. Section V introduces the proposed GLRT and studies its statistical performance. Numerical results obtained on simulated data are presented in Section VI. Finally, Section VII summarizes and concludes the paper.

II. NDN SECURITY BACKGROUND

In this section, we briefly introduce NDN, with a focus on the IFA and its recently proposed solutions.

A. Named Data Networking

ICN is a networking paradigm which is based on content objects. The novelty of ICN is its key concept of naming content objects, instead of naming hosts with IP addresses. Among ICN proposals, NDN is considered to be the most promising candidate to replace the current IP network. NDN uses a hierarchical naming scheme for content objects. Communications in NDN are performed by two types of packets: (1) *Interest* and (2) *Data*. A user sends an Interest packet to request some content and receives a Data packet containing the requested content in return. In NDN, a router has many faces - a generalization of interfaces in IP network. New router's components are also introduced in NDN. First, the

Content Store is an essential local cache that improves content delivery by storing recently requested content. Secondly, the *Forwarding Information Base* contains routing information for Interest packets. Finally, the *Pending Interest Table* (PIT) contains routing information for Data packets. More precisely, for each forwarded Interest, its incoming faces are saved in a PIT entry, so that the corresponding Data can be sent back to the user. For each received Data, the corresponding PIT entry will be removed.

B. Interest Flooding attack

It has been shown in [3] that the PIT can be overloaded by *Interest flooding attack* (IFA) - a variation of the Denial of Service (DoS) attack in NDN. The principle of IFA is to send a large amount of malicious Interest packets for non-existent contents. Such Interests cannot be resolved by any Data packet. Hence, the corresponding PIT entry cannot be removed. When the PIT is overloaded, new Interest packets cannot be handled and, thus, are dropped. This attack can have serious consequence on network for two reasons. First, it can cause large scale damage by targeting the network infrastructure. Secondly, Interests for non-existing content can be created easily.

C. Previously proposed solutions

Several solutions against IFA have been proposed so far [4]–[7]. Despite using different methods to weaken the attack’s damage, major part of proposed solutions rely on a detection phase, followed by a mitigation step. Previous detections are usually based on the packet-loss rate, since when the PIT is full, new incoming Interests are all dropped, increasing the loss-rate. However, those prior works present common limitations. First, those detectors are based on empirical experiences, with specific setup. Hence, their optimality in different conditions is questionable. Secondly, they are evaluated with easily detected cases (i.e. large attack payload to corrupt the PIT quickly). Therefore, their performance is unknown in realistic cases. Thirdly, those detectors are built on a limited exploitation of detection theory. Therefore, their detection threshold is not clearly defined and the setting in different routers with different traffic conditions is difficult. Finally, the underlying packet-loss rate model used in those detectors is simplistic. For instance, a constant packet-loss rate is assumed at all the routers in [5].

Our previous work [8] has proposed a Uniformly Most Powerful test against IFA for the case where packet-loss rate is perfectly known. In this paper, a parametric model for the evolution of loss-packet rate in IFA is introduced. Using the proposed model, we designed a practical GLRT against IFA in a scenario where the loss-packet rate is unknown in advance.

III. INTEREST FLOODING DETECTION PROBLEM DESCRIPTION

A. Definitions

In this paper, it is assumed that for each face of a NDN router, at instant t^{th} , the number of incoming Interest packets

and outgoing Data packets, denoted as i_t and d_t respectively, will be recorded. Ideally, each incoming Interest at a face should be resolved by one Data packet. However, in any type of networks, part of the packets could be lost. Hence, let $\ell_t = 1 - d_t/i_t$ be the measured packet-loss rate at instant t , i.e. the measured rate of unresolved Interests. The number of Interests i_t is drawn from a Poisson distribution - a usual model to represent the users’ behavior over the Internet [16]. In addition, following the model proposed in [4], [5], it is assumed that, at instant t , all Interests have the same probability of not being resolved, denoted as p_t . Under normal situation, such probability should corresponds to the expectation of packet-loss rate: $\mathbb{E}(\ell_t) = p_t$. It follows that the number of Data packets received d_t follows a binomial distribution $\mathcal{B}(i_t, 1 - p_t)$ with expectation $\mathbb{E}(d_t) = i_t(1 - p_t)$. For generality, it is assumed in this paper that the packet-loss rate ℓ is unknown and may change over time. However, under normal condition, its fluctuation should not be abrupt; we use this fact to build a model for the packet-loss rate of legitimate traffic. By contrast, when an IFA is started, a significant number of Interests for non-existing content are sent, resulting in an abrupt increase of the packet-loss rate ℓ .

Let us denote N_a the number of malicious Interests sent during the IFA by attacker-controlled hosts besides the legitimate Interests, denoted as i_t^* . Hence, the IFA can be characterized by an increase in the number of incoming Interests:

$$i_t = i_t^* + N_a. \quad (1)$$

It is important to note that distinguishing legitimate Interests i_t^* from the whole flow of Interest packets i_t is not possible. Moreover, because N_a additional Interests cannot be resolved, this will increase the overall packet-loss rate’s expectation as follows:

$$a = \mathbb{E}(\ell_t) - p_t = \frac{(1 - p_t)N_a}{i_t^* + N_a}. \quad (2)$$

The above relation comes from the fact that whether an IFA is currently happening or not, the expected number of Data packets received at a given face remains the same:

$$\begin{aligned} (1 - p_t)i_t^* &= \mathbb{E}(d_t) = (1 - p_t - a)(i_t^* + N_a), \\ \Leftrightarrow \frac{(1 - p_t)N_a}{i_t^* + N_a} &= a. \end{aligned}$$

B. Detection Problem Statement

Let us assume for the problem description that the packet-loss rate p_t is known. According to IFA’s presentation in Section III-A, the detection problem against IFA consists in choosing between two hypotheses: \mathcal{H}_0 : “the number of Interests sent i_t and Data packets received d_t are consistent with what is expected from p_t ” and \mathcal{H}_1 : “the number of Interest packets sent i_t is significantly higher than what is expected from d_t and p_t ”. Those two can be written formally as:

$$\begin{cases} \mathcal{H}_0 : d_t \sim \mathcal{B}(i_t, 1 - p_t), \\ \mathcal{H}_1 : d_t \sim \mathcal{B}(i_t - N_a, 1 - p_t), N_a > 0. \end{cases} \quad (3)$$

Formally, a statistical test is a mapping $\delta : \mathbb{R} \mapsto \{\mathcal{H}_0; \mathcal{H}_1\}$ such that hypothesis \mathcal{H}_i , $i \in \{0, 1\}$ is accepted if $\delta(x) = \mathcal{H}_i$ (see [9] for a thorough introduction to hypothesis testing theory). This paper focuses on the Neyman-Pearson bi-criteria approach that simultaneously aims at guaranteeing a false-alarm probability while maximizing the power function (or correct-detection probability). For the sake of definition, let $\mathbb{P}_i(E)$, $i \in \{0, 1\}$ be the probability of event E under hypothesis \mathcal{H}_i . For a prescribed false-alarm probability α_0 , the Neyman-Pearson approach aims at finding a test δ whose false alarm probability is upper bounded by α_0 . Hence, let:

$$\mathcal{K}_{\alpha_0} = \{\delta : \mathbb{P}_0[\delta(\ell_t) = \mathcal{H}_1] \leq \alpha_0\}, \quad (4)$$

be the class of all tests with a false-alarm probability upper bounded by α_0 . The goal is to find in \mathcal{K}_{α_0} a test maximizing the power function, which is formally defined by the correct detection probability:

$$\beta_\delta = \mathbb{P}_1[\delta(\ell_t) = \mathcal{H}_1], \quad (5)$$

which is equivalent to minimizing the missed-detection probability $\alpha_1(\delta) = 1 - \beta_\delta$.

The hypotheses as formulated in Equation (3) highlight the main difficulties of the present testing problem. First, we note that the IFA implies a change on both expectation and variance of the measured loss-packet rate ℓ_t . Second, the parameters of attack payload N_a or a (Equations (1) - (2)) are unknown. Ideally, the test δ should maximize the power function β_δ , regardless of the attack payload. Such a test that maximizes the power function uniformly with respect to the attack payload a is a Uniformly Most Powerful (UMP) test. Unfortunately, UMP test scarcely ever exists. Third, the undoubtedly greatest difficulty is that the expected loss-packet rate is unknown in practice. This problem is addressed in the Section V.

IV. OPTIMAL LIKELIHOOD RATIO TEST FOR KNOWN LOSS RATE

In this section, it is assumed that the expected packet-loss rate p_t is known. As a result, This allows designing the theoretical optimal Likelihood Ratio Test (LRT) and also permits the assessment of its statistical performance. Since p_t is already known for each instant t , the LRT does not necessarily require measured values from previous instants. Therefore, i_t and d_t are adequate for the test's input. The case of unknown packet-loss rate ℓ will be addressed in the next section.

Since the binomial law belongs to the family of the exponential distribution, there exists a UMP test that is given by the following decision rule, see [9, Corollary 3.4.1]:

$$\delta^*(d_t) = \begin{cases} \mathcal{H}_0 & \text{if } d_t \geq h, \\ \mathcal{H}_1 & \text{if } d_t < h, \end{cases} \quad (6)$$

where h is a threshold such that $\delta^* \in \mathcal{K}_{\alpha_0}$ (4). However, we note that evaluating the statistical properties of this test, especially the false alarm and the power function, may be difficult. To simplify this task, it is proposed in this paper

to apply the central limit theorem (CLT) [9, theorem 11.2.5], assuming that the number of Interest sent i_t is large, which is a very usual case for a router face. Hence, for legitimate traffic, the number of Data packets received can be modeled as:

$$d_t \rightsquigarrow \mathcal{N}(i_t(1-p_t), i_t p_t(1-p_t)), \quad (7)$$

where \rightsquigarrow represents the convergence in distribution as i_t tends to infinity. For clarity, it is also proposed to replace the decision rule (6) by introducing the residual $r_t = \ell_t - p_t$. Hence, under normal situation \mathcal{H}_0 :

$$r_t = \left(1 - \frac{d_t}{i_t}\right) - p_t \rightsquigarrow \mathcal{N}\left(0, \frac{p_t(1-p_t)}{i_t}\right). \quad (8)$$

On the opposite, when an IFA is started, the residual tends to:

$$r_t \rightsquigarrow \mathcal{N}\left(a, \frac{p_t(1-p_t)}{i_t} - \frac{N_a p_t(1-p_t)}{i_t^2}\right). \quad (9)$$

For simplicity and clarity, it is proposed to denote σ_t^2 the variance under \mathcal{H}_0 and σ_a^2 the decrease of variance due to the IFA:

$$\sigma_t^2 = \frac{p_t(1-p_t)}{i_t}, \quad \sigma_a^2 = \frac{N_a p_t(1-p_t)}{i_t^2} = \frac{a p_t}{i_t}. \quad (10)$$

We can note that the decrease of variance is due to the increase in number of Interest packets sent during the attack, i.e. from $i_t = i_t^*$ in Equation (8) to $i_t = i_t^* + N_a$ in Equation (9) while the number of receive Data packet does not change. From Equations (8) - (9), the testing problem (3) can reformulated as:

$$r_t \sim \begin{cases} \mathcal{N}(0, \sigma_t^2) & \text{under } \mathcal{H}_0, \\ \mathcal{N}(a, \sigma_t^2 - \sigma_a^2) & \text{under } \mathcal{H}_1. \end{cases} \quad (11)$$

Equation (11) shows that the decrease of the packet-loss rate's expectation a characterizes the distribution of the residual under \mathcal{H}_1 , hence it is used in the remaining of this paper to quantify the payload of the attack.

One can note that the function $f : d_t \rightarrow R(d_t) = r_t$ is strictly decreasing. Hence, the test $\delta^*(d_t)$ is equivalent to the following test:

$$\delta^*(r_t) = \begin{cases} \mathcal{H}_0 & \text{if } r_t \leq \tau^* = R(h), \\ \mathcal{H}_1 & \text{if } r_t > \tau^* = R(h). \end{cases} \quad (12)$$

As previously discussed, the application of the CLT (8) allows establishing the statistical properties of the optimal UMP test, presented in the following Proposition 1:

Proposition 1. *Assuming that the number of Interest i_t tends to infinity, for any prescribed false-alarm probability α_0 , the decision threshold, τ^* , given by:*

$$\tau^*(\alpha_0) = \Phi^{-1}(1 - \alpha_0) \sigma_t, \quad (13)$$

guarantees that the test δ^ (12) is in \mathcal{K}_{α_0} . Here Φ and Φ^{-1} are the standard normal cumulative distribution function and its inverse function. Using the decision threshold given in (13) the power function of the UMP test δ^* (12) is given by:*

$$\beta_{\delta^*}(a) = 1 - \Phi\left(\frac{\sigma_t \Phi^{-1}(1 - \alpha_0) - a}{\sqrt{\sigma_t^2 - \sigma_a^2}}\right). \quad (14)$$

V. PROPOSED GENERALIZED LIKELIHOOD RATIO

A. Packet-loss Rate Model

This sections addresses the case where the expected packet-loss rate p_t is unknown. In order to enhance the detection, N last measurements of packets-loss rate $\ell = (\ell_{T-N+1}, \dots, \ell_T)$ will be taken into account. Since the fluctuation of the packet-loss rate is limited and smooth [10] [11], its expectation can be modeled by a polynomial: $\mathbf{p} = \mathbf{H}\mathbf{x}$, where \mathbf{H} is a matrix of size $N \times q$ with $\mathbf{H}_{tj} = t^{j-1}$, $t \in \{T-N+1, \dots, T\}$, $j \in \{0, \dots, q-1\}$ and $\mathbf{x} = (x_0, \dots, x_{q-1})$ is the vector of the q coefficients of the polynomial. Such a model has been widely used in signal processing, see [12]–[14] for applications in Internet traffic modeling and image processing.

Assuming that packet-loss rate measurements ℓ are independent, the CLT allows modeling those observations as follows:

$$\ell \rightsquigarrow \mathcal{N}(\mathbf{H}\mathbf{x}, \Sigma_0), \quad (15)$$

where Σ_0 is a diagonal covariance matrix whose element is $\frac{p_t(1-p_t)}{i_t}$, $t \in \{T-N+1, \dots, T\}$.

When an attack is started at instant T^{th} , the last samples d_T and i_T will be affected by the increase in number of Interest packets (1) sent and packet-loss rate ℓ_T (2). Hence under hypothesis \mathcal{H}_1 , as i_T tends to infinity, the packet-loss rate will tends to the following model:

$$\ell \rightsquigarrow \mathcal{N}(\mathbf{H}\mathbf{x} - a\mathbf{v}_a, \Sigma_0 - \Sigma_a), \quad (16)$$

where Σ_a , as in Equation (10), represents the decrease of variance due to the IFA that only affects the corrupted samples, and \mathbf{v}_a represents the change in loss-packet rate due to the attack, for instance, $\mathbf{v}_a = (0, 0, \dots, 0, 1)^T$ when only the very last sample is corrupted.

In this paper, it is proposed to use the least square method to estimate the expectation of the packet-loss rate \mathbf{p} :

$$\tilde{\mathbf{p}} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\ell,$$

and consequently the estimated residuals \mathbf{r} are defined, as in Equation (8), as:

$$\tilde{\mathbf{r}} = \ell - \tilde{\mathbf{p}} = \mathbf{H}^\perp\ell, \quad (17)$$

where $\mathbf{H}^\perp = \mathbf{I}_N - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$, with \mathbf{I}_N the identity matrix of size N , represents the projection onto the orthogonal complement of the subspace spanned by the columns of \mathbf{H} .

B. Proposed Test and Study of its Properties

From the model of the packet-loss rate under each hypothesis (15) - (16), one can note that the testing problem with unknown packet-loss rate can be formulated as a choice between the following hypotheses:

$$\begin{cases} \mathcal{H}_0 : \tilde{\mathbf{r}} \sim \mathcal{N}(\mathbf{0}, \mathbf{H}^\perp\Sigma_0\mathbf{H}^{\perp T}), \\ \mathcal{H}_1 : \tilde{\mathbf{r}} \sim \mathcal{N}(a\tilde{\mathbf{v}}_a, \mathbf{H}^\perp\Sigma_0\mathbf{H}^{\perp T} - \mathbf{H}^\perp\Sigma_a\mathbf{H}^{\perp T}), \end{cases} \quad (18)$$

with $\tilde{\mathbf{v}}_a = \mathbf{H}^\perp\mathbf{v}_a$ the footprint of the IFA after estimating and then removing the expected loss-packet rate (17). Here it can be noted that, as previously discussed in Section IV, the

IFA impacts both the expectation and the covariance of the residuals.

Obviously, designing an optimal test for the hypothesis testing problem (18) is challenging. In this paper, it is proposed to apply the UMP test designed in the case of a known packet-loss rate by replacing the residuals by the estimated ones, from Equation (17). This leads to the Generalized LRT that is defined by:

$$\tilde{\delta}(\tilde{\mathbf{r}}) = \begin{cases} \mathcal{H}_0 & \text{if } \tilde{\mathbf{v}}_a^T\tilde{\mathbf{r}} \leq \tilde{\tau}, \\ \mathcal{H}_1 & \text{if } \tilde{\mathbf{v}}_a^T\tilde{\mathbf{r}} > \tilde{\tau}. \end{cases} \quad (19)$$

The very interesting results this paper proposes is that, because it is possible to establish the statistical distribution of the Generalized Likelihood Ratio $\tilde{\delta}(\tilde{\mathbf{r}})$, one can establish analytically the properties of the proposed test.

From the distribution of the residuals $\tilde{\mathbf{r}}$, Equation (18), it is straightforward that:

$$\tilde{\mathbf{v}}_a^T\tilde{\mathbf{r}} \rightsquigarrow \begin{cases} \mathcal{N}(\mathbf{0}, s_0^2) & \text{under } \mathcal{H}_0, \\ \mathcal{N}(a\|\tilde{\mathbf{v}}_a\|_2^2, s_0^2 - s_a^2) & \text{under } \mathcal{H}_1. \end{cases} \quad (20)$$

where for clarity and simplicity the variance of the GLR under \mathcal{H}_0 is given by:

$$s_0^2 = \mathbf{v}_a^T\mathbf{H}^\perp\Sigma_0\mathbf{H}^{\perp T}\mathbf{v}_a,$$

and, similarly, the decrease of variance of the GLR under \mathcal{H}_1 is defined by:

$$s_a^2 = \mathbf{v}_a^T\mathbf{H}^\perp\Sigma_a\mathbf{H}^{\perp T}\mathbf{v}_a.$$

Similarly to Proposition 1, we can establish the decision threshold and the power function of the proposed GLRT as following:

Proposition 2. *Assuming that the number of Interest i_t tends to infinity, for any prescribed false-alarm probability α_0 , the decision threshold $\tilde{\tau}$ given by:*

$$\tilde{\tau} = \Phi^{-1}(1 - \alpha_0) s_0, \quad (21)$$

guarantees that the test $\tilde{\delta}$ (19) is in \mathcal{K}_{α_0} . Using the decision threshold given in (21), the power function of the UMP test (19) is given by:

$$\beta_{\delta^*}(a) = 1 - \Phi\left(\frac{s_0\Phi^{-1}(1 - \alpha_0) - a\|\tilde{\mathbf{v}}_a\|_2^2}{\sqrt{s_0^2 - s_a^2}}\right). \quad (22)$$

From the power function (22), one can note that the loss of optimality of the proposed GLRT is mainly caused by the factor $\|\tilde{\mathbf{v}}_a\|_2^2$. This is explained by the fact that a non-negligible proportion of the packet-loss rate's change due to IFA will be modeled as part of the regular change due to legitimate traffic.

VI. NUMERICAL RESULTS

A. Scenario setup

In the present paper, two sets of numerical results are presented. First, results obtained on data simulated under *Matlab* are presented to verify the sharpness of the theoretical findings. Then, *ndnSIM* - an open source NDN simulator, provided by

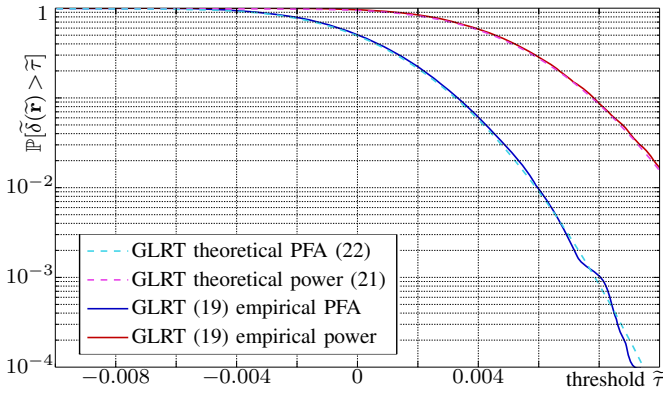


Fig. 1: Comparison between proposed GLRT theoretical false-alarm probability (PFA) and detection power $\beta_{\delta^*}(a)$, see Equation (22), and empirical ones. The false-alarm probability and power are plotted as a function of decision threshold $\tilde{\tau}$.

the NDN project - is used to generate more realistic data. Indeed, ndnSIM faithfully implements the components of a NDN network which allows us to consider every aspect of the network [15]. Besides, in order to compare the performance of our approach to the existing ones, we reuse one of the topologies from [4] - a binary tree with 8 hosts, intermediate routers and one content provider for our evaluation. The experimental settings are also referred to our prior work [8] that uses the same topology.

In all of our simulations, the actual number of Interests sent is generated from a Poisson distribution whose mean value is drawn from a uniform random variable. In addition, the actual packet-loss rate follows an auto-regressive (AR) model. Such model has been extensively used to model both users' requests evolution and packet-loss rate in computer network [11], [17] and can be easily implemented in ndnSIM. More precisely, the packet-loss rate is initialized at $p_0 = 0.05$, then following expectations of packet-loss rates are given by $p_t = p_{t-1} + u$ with u the realization of a uniformly distributed random variable. To avoid computational problem, the sign of u is flipped if $p_t < 0$. Several values for those parameters have been tested and the obtained results show similar trends.

Finally, note that for the proposed GLRT, a set of 50 samples was used and the degree of the polynomial is $q - 1 = 4$, hence the matrix \mathbf{H} has the size 50×5 . In all the figures, except Figure 4, it is considered that the quickest detection is desired, hence it is aimed at detecting if only the last sample is corrupted. In such a case the footprint of IFA on the packet-loss rate is characterized by \mathbf{v}_a that has non zeros only on its last element, which give a footprint after packet-loss rate estimation $\tilde{\mathbf{v}}_a$ with $\|\tilde{\mathbf{v}}_a\|_2^2 \approx 0.6$.

B. Numerical results on simulated data

Because one of the main goal of the present paper is to establish the statistical properties of the proposed GLRT, Figure 1 shows a comparison between the theoretical probabilities of false-alarm and detection power, given in Proposition 2, and the empirical ones. Even for threshold that corresponds

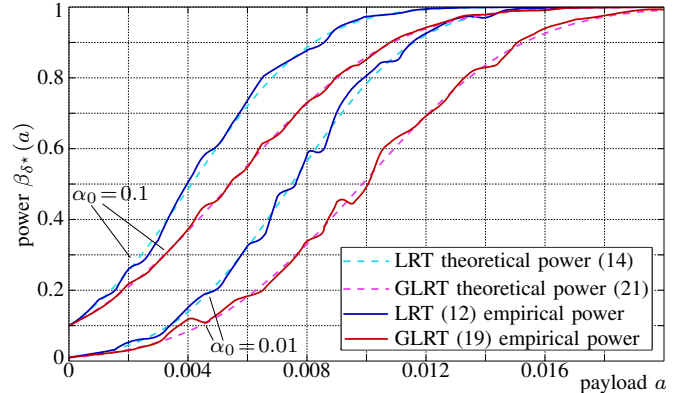


Fig. 2: Comparison between empirical detection power and theoretical power function for both optimal LRT and the proposed GLRT. The power function is plotted as a function of the strength of the anomaly $a \in [0, 0.02]$.

to probabilities as small as 10^{-3} , the empirical results match well the theoretically established ones. This observation is important as this guarantees a prescribed false-alarm probability in a practical situation. This also shows the sharpness of the theoretical findings and relevance of the proposed model.

Then Figure 2 compares the theoretical and empirical power as a function of the IFA payload a for both optimal LRT and proposed GLRT. We also note that the power is computed with two prescribed false-alarm rates $\alpha_0 = 0.01$ and $\alpha_0 = 0.1$. Figure 2 again shows the relevance of the theoretical findings since empirical power functions match the theoretical ones. However, for low false-alarm probability as $\alpha_0 = 0.01$, the number of required samples is very large, hence, empirical results are slightly less accurate.

C. Numerical results on ndnSIM data

It is hardly possible to obtain real data from NDN routers, as NDN is not yet deployed at large scale. Hence, it is proposed to verify the relevance of the proposed approach on data as close to reality as possible by using ndnSIM. Figure 3 presents, similarly to Figure 1, the comparison between the proposed GLRT theoretical and empirical false-alarm probability as a function of detection threshold $\tilde{\tau}$. Because the actual values of packet-loss rates are unknown, the optimal LRT cannot be included in this comparison. We note from Figure 1 that the number of samples is much smaller since running ndnSIM is time consuming. However, even with this limited number of samples, the empirical probability of false alarm matches the theoretical one. This results is very important as it shows that the proposed approach remains accurate with data "as close as possible" from real ones.

As discussed in Section VI-A, previous figures focus on the case in which only one sample is corrupted by the IFA. Hence, Figure 4 shows a comparison between the theoretical and the empirical power of the proposed GLRT for three number of corrupted sample, denoted M , 1, 3 and 7. As one would expect, the power increases with the number of corrupted samples. This result emphasizes that the proposed method can

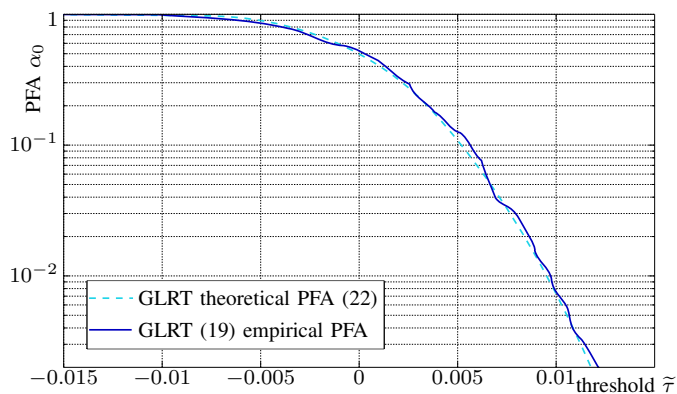


Fig. 3: Comparison between empirical and theoretical false alarm probability for the proposed GLRT, as the function of decision threshold $\tilde{\tau}$.

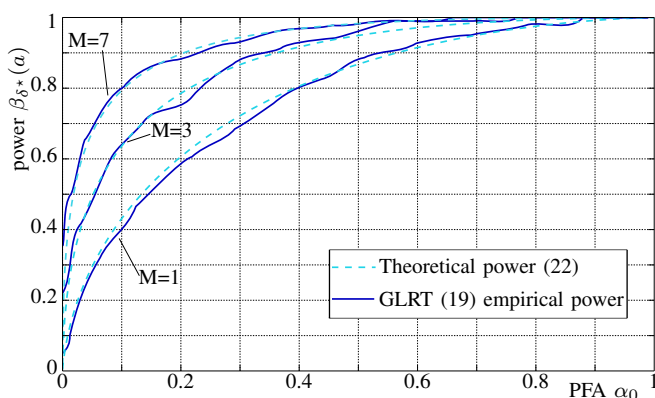


Fig. 4: Receiver Operational Characteristic (ROC) curves for the proposed GLRT with different number of samples corrupted.

be adapted to focus on the quickest detection, hence aiming at detecting only if the last sample is corrupted at a cost of lower detection accuracy. On the other hand, it is also possible to increase the detection delay, hence focusing on the detection of several last samples corrupted by the IFA, to ensure a higher detection accuracy.

VII. CONCLUSION

In this paper, a parametric model for the evolution of loss-packet rate in IFA has been proposed. Moreover, a practical GLRT is designed by applying this model in a scenario where the loss-packet rate is unknown in advance. The proposed test is upper bounded by an optimal LRT designed for the theoretical case with known loss rate. The relevance of our proposed model has been proved with both simulated data from MATLAB and from ndnSIM. The proposed method can be adapted to focus on either to quickly detect the very last sample with cost of lower detection accuracy, or to achieve better detection accuracy in difficult situation with longer delay.

ACKNOWLEDGEMENTS

This work is funded by the French National Research Agency (ANR), DOCTOR project <ANR-14-CE28-000> and by the French Systematic ICT cluster.

REFERENCES

- [1] G. Xylomenos & al., "A survey of information-centric networking research," *Communications Surveys & Tutorials, IEEE*, vol. 16, no. 2, pp. 1024–1049, 2014.
- [2] L. Zhang & al., "Named data networking," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 3, pp. 66–73, 2014.
- [3] P. Gasti & al., "Dos and ddos in named data networking," in *Computer Communications and Networks (ICCCN), IEEE 22nd International Conference on*. 2013, pp. 1–7.
- [4] A. Afanasyev & al., "Interest flooding attack and countermeasures in named data networking," in *IFIP Networking Conference, 2013*. IEEE, 2013, pp. 1–9.
- [5] A. Compagno & al., "Poseidon: Mitigating interest flooding ddos attacks in named data networking," in *Local Computer Networks (LCN), IEEE 38th Conference on*. IEEE, 2013, pp. 630–638.
- [6] H. Dai & al., "Mitigate ddos attacks in ndn by interest traceback," in *Computer Communications Workshops (INFOCOM WKSHPS), 2013 IEEE Conference on*. IEEE, 2013, pp. 381–386.
- [7] K. Wang & al., "Detecting and mitigating interest flooding attacks in content-centric network," *Security and Communication Networks*, vol. 7, no. 4, pp. 685–699, 2014.
- [8] N. T. Nguyen, R. Cogranne, and G. Doyen, "An optimal statistical test for robust detection against interest flooding attacks in ccn," in *Integrated Network Management (IM), IFIP/IEEE International Symposium on*. IFIP/IEEE.
- [9] E. Lehmann and J. Romano, *Testing Statistical Hypotheses, Second Edition*, 3rd ed. Springer, 2005.
- [10] J.-C. Bolot, "Characterizing end-to-end packet delay and loss in the internet," *J. High Speed Networks*, vol. 2, no. 3, pp. 305–323, 1993.
- [11] E. Altman, K. Avrachenkov, and C. Barakat, "A stochastic model of tcp/ip with stationary random losses," *Networking, IEEE/ACM Transactions on*, vol. 13, no. 2, pp. 356–369, 2005.
- [12] H. Yin & al., "Network traffic prediction based on a new time series model," *International Journal of Communication Systems*, vol. 18, no. 8, pp. 711–729, 2005.
- [13] R. Cogranne and F. Retraint, "An asymptotically uniformly most powerful test for LSB matching detection," *Information Forensics and Security, IEEE Transactions on*, vol. 8, no. 3, pp. 464–476, March 2013.
- [14] R. Cogranne and F. Retraint, "Detection of defects in radiographic images using an adaptive parametric model," *Signal Processing*, vol. 96, Part B, pp. 173 – 189, March 2014.
- [15] A. Afanasyev & al., "ndnsim: Ndn simulator for ns-3," *University of California, Los Angeles, Tech. Rep*, 2012.
- [16] V. S. Frost and B. Melamed, "Traffic modeling for telecommunications networks," *IEEE Communications Magazine*, vol. 32, no. 3, pp. 70–81, 1994.
- [17] S. Basu, A. Mukherjee, and S. Klivansky, "Time series models for internet traffic," in *INFOCOM'96, IEEE International Conference on Computer Communications*, vol. 2, 1996, pp. 611–620.