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

Ngoc Tan 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 us to provide 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 that faces the future Internet requirements. Such a new network design have been proposed in *Information Centric Network* (ICN) [1] architectures. Among ICN proposals, *Named Data Networking* (NDN) [2] is a promising future network. This architecture draws a lot of attentions from 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, that is without much 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 on 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 us to guarantee a prescribed false-alarm probability and to compute 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 consider 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 type 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 send back to the user. For every 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 Interest packets for non-existent contents. Such Interests can not be resolved by any Data packet. Hence, the corresponding PIT entry can not be removed. When the PIT is overloaded, new Interest packets can not be handled and, thus, are dropped. This attack is highly risky 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 requests 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 for 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

We consider in this paper that all NDN routers can record periodically, for a given face, the number of incoming Interest packets, denoted $\mathbf{i} = (i_1, \dots i_T)$, and the number of outgoing Data packets, denoted $\mathbf{d} = (d_1 \dots d_T)$. Ideally, for each Interest sent, a Data packet should be returned. However, in any type of networks, part of the packets can be lost. Hence, let the packet-loss rate be defined as $\boldsymbol{\ell} = (\ell_1, \dots, \ell_T), \ell_t = 1 - d_t/i_t$, which is the measured rate of Interest packets that are not resolved. The number of Interest 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 Interest packets have the same probability of not being resolved, denoted p_t which 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 $\boldsymbol{\ell}$ is unknown and changes in time. However, the evolution of this packet-loss rate is usually not 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 rather significant number of Interest packets are sent with non-existing content which implies an abrupt increase in the packet-loss rate $\boldsymbol{\ell}$.

Let us denote N_a the number of Interest packets for non-existing content name sent by the attacker-controlled host during the IFA. Those Interests are send besides the legitimate traffic which remains unchanged. Hence, the IFA can be characterized by an increase in the number of incoming Interests:

$$i_t = i_t^\star + N_a, \tag{1}$$

where $i_t^\star$ represents the number of legitimate Interests and N_a the added number of Interests for non-existing content due to the IFA. It is important to note that distinguishing legitimate Interests $i_t^\star$ from the whole flow of Interest packets i_t is not possible. Moreover, because the N_a additional Interest packets cannot be resolved, this will increase the overall measured packet-loss rate as follows:

$$a = p_t - p_t^\star = \frac{(1 - p_t^\star)N_a}{i_t^\star + N_a}. \tag{2}$$

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:

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

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:

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

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 such as its false alarm probability is upper bounded by α_0. Hence, let:

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

be the class of all tests with a false-alarm probability upper bounded by α_0. In the class $\mathcal{K}_{\alpha_0}$ it is aimed at finding a test that maximizes the power function, formally defined by the correct detection probability:

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

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

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 $\beta_\delta(a)$ 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, such 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 p_t rate is known. This allows us to design the theoretical optimal Likelihood Ratio Test (LRT) and also permits the assessing of its statistical performance. The next Section V will address the case of unknown packet-loss rate $\boldsymbol{\ell}$.

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]:

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

where h is a threshold such that $\delta^\star \in \mathcal{K}_{\alpha_0}$ (4). However, we note that assessing the statistical properties of this test, especially the false alarm and the powerfunction, 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:

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

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 defined by:

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

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

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

For simplicity and clarity, it is proposed to denote σ_t^2 the variance under the assumption of legitimate traffic only and σ_a^2 the decrease of variance due to the IFA:

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

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

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

Equation (11) shows that the decrease of the packet-loss rate 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 $f : d_t \to R(d_t) = r_t$ is strictly decreasing, hence, the test $\delta^\star(d_t)$ is equivalent to the following test:

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

As previously discussed, the application of the CLT (8) allows to establish 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, $\tau^\star$, given by:*

$$\tau^\star(\alpha_0) = \Phi^{-1}(1-\alpha_0)\,\sigma_t. \tag{13}$$

guarantees that the test $\delta^\star$ (12) is in $\mathcal{K}_{\alpha_0}$. Here Φ and Φ^{-1} are the standard normal cumulative distribution function and its inverse. Using the decision threshold given in (13) the power function of the UMP test $\delta^\star$ (12) is given by:

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

V. Proposed Generalized Likelihood Ratio

A. Packet-loss Rate Model

Let us now study the case where the packet-loss rate ℓ_t, $t = \{1,\dots,T\}$ is unknown. To this end, the N last measurements of packets-loss rate are gathered $\boldsymbol{\ell} = (\ell_{T-N+1},\dots,\ell_T)$. Since the evolution of the packet-loss rate is limited and smooth [10] [11] its expectation is modeled using a polynomial $\mathbf{p} = \mathbf{H}\mathbf{x}$ where the column of matrix $\mathbf{H}$, of size $N \times q$, spans the basis of polynomial $t,\dots,t^{q-1}$ and $\mathbf{x}$ is the vector of the q coefficients. 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 are independent, the central limit theorem allows the modeling of those observations $\boldsymbol{\ell}$ as follows:

$$\boldsymbol{\ell} \rightsquigarrow \mathcal{N}(\mathbf{H}\mathbf{x}\,,\,\boldsymbol{\Sigma}_0), \tag{15}$$

with the covariance matrix $\boldsymbol{\Sigma}_0$ a diagonal matrix whose elements are $\frac{p_n(1-p_n)}{i_n}$, $n \in \{T-N+1,\dots,T\}$.

When an attack is started at time T, 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 this hypothesis $\mathcal{H}_1$ the loss packet rate will tends, as i_T tends to infinity, to the following model:

$$\boldsymbol{\ell} \rightsquigarrow \mathcal{N}(\mathbf{H}\mathbf{x} - a\mathbf{v}_a\,,\,\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_a), \tag{16}$$

where $\boldsymbol{\Sigma}_a$, 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 $\mathbf{p}$:

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

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

$$\widetilde{\mathbf{r}} = \mathbf{p} - \widetilde{\mathbf{p}} = \mathbf{H}^\perp\mathbf{p}, \tag{17}$$

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:

$$\begin{cases} \mathcal{H}_0 = \left\{\widetilde{\mathbf{r}} \sim \mathcal{N}\left(\mathbf{0}\,,\,\mathbf{H}^\perp\boldsymbol{\Sigma}_0{\mathbf{H}^\perp}^T\right)\right\}, \\ \mathcal{H}_1 = \left\{\widetilde{\mathbf{r}} \sim \mathcal{N}\left(a\widetilde{\mathbf{v}}_a\,,\,\mathbf{H}^\perp\boldsymbol{\Sigma}_0{\mathbf{H}^\perp}^T - \mathbf{H}^\perp\boldsymbol{\Sigma}_a{\mathbf{H}^\perp}^T\right)\right\}, \end{cases} \tag{18}$$

with $\widetilde{\mathbf{v}}_a = \mathbf{H}^\perp\mathbf{v}_a$ 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:

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

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

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

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

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

$$s_0^2 = \mathbf{v}_a^T\mathbf{H}^\perp\boldsymbol{\Sigma}_0{\mathbf{H}^\perp}^T\mathbf{v}_a,$$

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

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

We are now ready to established the statistical properties of the proposed GLRT, in the following Proposition 2, that analogously to Proposition 1 establish the decision threshold and the power function.

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

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

guarantees that the test $\widetilde{\delta}$ 19 is in $\mathcal{K}_{\alpha_0}$. Using the decision threshold given in (21) the power function of the UMP test (19) is given by:

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

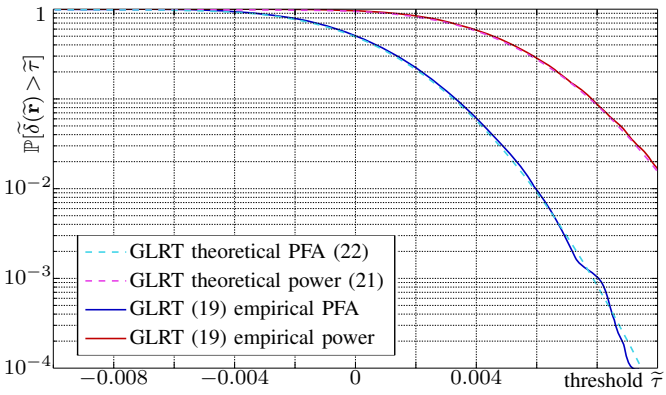

Fig. 1: Comparison between proposed GLRT theoretical false-alarm probability (PFA) and detection power $\beta_{\delta^\star}(a)$, see Equation (22), and empirical ones. The false-alarm probability and power are plotted as a function of decision threshold $\widetilde{\tau}$.

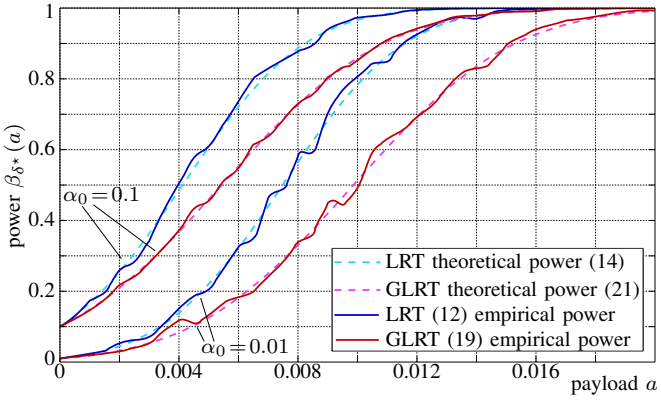

Fig. 2: Comparison between empirical detection power and theoretical power function for both optimal LRT and the proposed GLRT. The power function is plotted as a function of the strenght of the anomaly $a \in [0, 0.02]$.

From the power function (22) one can note that the loss of optimality of the proposed GLRT is mainly caused by the factor $\|\widetilde{\mathbf{v}}_a\|_2^2$. This is explained by the fact that a non-negligible proportion of the packet-loss rate changes 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 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 mean number of interest packets sent is drawn from a uniform random variable and this parameter is used to generate the actual number of interest sent, which is drawn from a Poisson distribution.

In addition, in our simulation the actual packet-loss rate follows an auto-regressive (AR) model. This model has been chosen because such a model can easily be implemented in ndnSIM and because it has been extensively used to model both users' requests evolution and packet-loss rate in computer network [11], [17]. 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 $\mathbf{A}$ that has non zeros only on its last element, which give a footprint after packet-loss rate estimation $\widetilde{\mathbf{v}}_a$ with $\|\widetilde{\mathbf{v}}_a\|_2^2 \approx 0.6$.

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. We note that even for threshold that corresponds to probabilities as small as 10^{-3}, the empirical results match well the theoretically established one. This observation is important as 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. We note however, that 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 it is not yet deployed at large scale, it is proposed to verify the relevance of the proposed approach on data that as close as possible from reality using ndnSIM. To this end, Figure 3 presents, similarly to Figure 1, the comparison between the

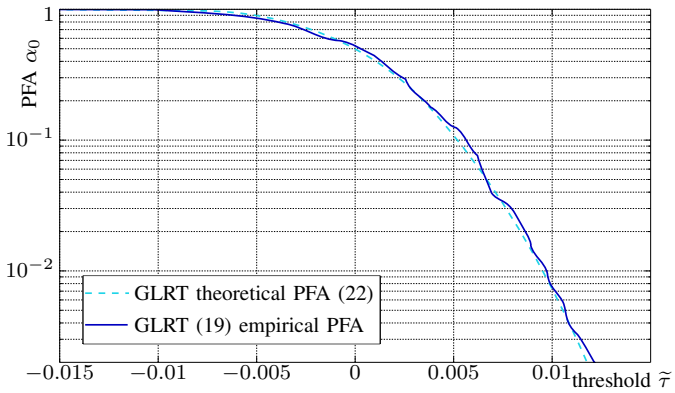

Fig. 3: Comparison between empirical detection power and theoretical power function for both optimal LRT and the proposed GLRT. The power function is plotted as a function of the strenght of the anomaly $a \in [0, 0.02]$.

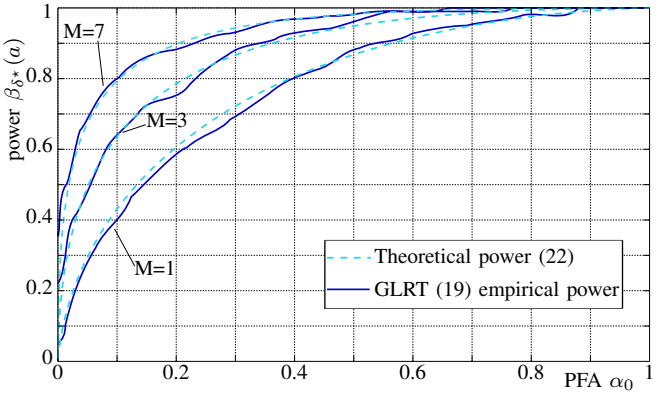

Fig. 4: Receiver Operational Characteristic (ROC) curves for the proposed GLRT with different number of samples corrupted.

proposed GLRT theoretical and empirical false-alarm probability as a function of detection threshold $\widetilde{\tau}$. 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 because running ndnSIM is time consuming, however, even with this limited number of samples, the empirical probability of false alarm match the theoretical one. This results is very important as its 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. In Figure 4 it is proposed to present the evolution of the proposed GLRT power as a function of number of corrupted samples. 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 interesting results emphasizes that the proposed method can 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 opposite, it is also possible to increase the detection delay, hence focusing on the detection of several lasts 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.