

# Efficient Computer Network Anomaly Detection by Changepoint Detection Methods

Alexander G. Tartakovsky, *Senior Member, IEEE*, Aleksey S. Polunchenko, and Grigory Sokolov

**Abstract**—We consider the problem of efficient on-line anomaly detection in computer network traffic. The problem is approached statistically, as that of sequential (quickest) changepoint detection. A multi-cyclic setting of quickest change detection is a natural fit for this problem. We propose a novel score-based multi-cyclic detection algorithm. The algorithm is based on the so-called Shiryaev–Roberts procedure. This procedure is as easy to employ in practice and as computationally inexpensive as the popular Cumulative Sum chart and the Exponentially Weighted Moving Average scheme. The likelihood ratio based Shiryaev–Roberts procedure has appealing optimality properties, particularly it is exactly optimal in a multi-cyclic setting geared to detect a change occurring at a far time horizon. It is therefore expected that an intrusion detection algorithm based on the Shiryaev–Roberts procedure will perform better than other detection schemes. This is confirmed experimentally for real traces. We also discuss the possibility of complementing our anomaly detection algorithm with a spectral-signature intrusion detection system with false alarm filtering and true attack confirmation capability, so as to obtain a synergistic system.

**Index Terms**—Changepoint detection, cybersecurity, CUSUM, distributed denial-of-service attacks, intrusion detection, network traffic, Shiryaev–Roberts Procedure, TCP SYN attacks.

## I. INTRODUCTION

THE Internet has never been a safe place and designing automated and efficient techniques for rapid detection of computer network anomalies (e.g., due to intrusions) never ceased to be a topical problem in cybersecurity [1]. Many existing anomaly-based Intrusion Detection Systems (IDS-s) operate by applying the machinery of statistics to comb through the passing traffic looking for a deviation from the traffic’s normal profile [2]–[6]. By way of example, the Sequential Probability Ratio Test (SPRT) [7], the Cumulative Sum (CUSUM)

chart [8], and the Exponentially Weighted Moving Average (EWMA) inspection scheme [9] are the *de facto* “workhorse” of the community. The CUSUM and EWMA methods come from the area of sequential changepoint detection, a branch of statistics concerned with the design and analysis of a *fastest* way to detect a change (i.e., an anomaly) in the state of a phenomenon (time process) of interest [10], [11].

Yet another changepoint detector popular in statistics is the Shiryaev–Roberts (SR) procedure [12]–[14]. Though practically unknown in the cybersecurity community, the SR procedure is as computationally simple as the CUSUM chart or the EWMA scheme. However, unlike the latter two, the SR procedure is also the best one can do (i.e., exactly optimal) in a certain multi-cyclic setting [15], a natural fit in the computer network anomaly detection context. The aim of this work is to offer a novel multi-cyclic anomaly detector using the SR procedure as the prototype. Due to the exact multi-cyclic optimality of the SR procedure, the proposed algorithm is expected to outperform other detection schemes, in particular the multi-cyclic CUSUM procedure. We confirm this experimentally using real data.

The remainder of the paper is organized as follows. Section II provides an introduction to the subject of changepoint detection. In Section III, we present our anomaly detection algorithm. In Section IV, we illustrate our algorithm at work. In Section V, we comment on how to improve the performance of the algorithm. Lastly, Section VI draws the conclusions.

## II. QUICKEST CHANGEPOINT DETECTION

Quickest changepoint detection is a study of techniques to detect a change (“disorder”) in the state of a time process, usually from “normal” to “abnormal”; inference about the process’ current state is made from a series of quantitative random observations (e.g., measurements corrupted by noise). The sequential setting assumes the series is amassed one at a time, and so long as the recorded data behavior suggests the process is in its “normal” state it is let to continue. However, if the observations hint that the process’ state may have switched to “abnormal”, ones aim is to detect the true change as quickly as possible for a given risk associated with false alarms, so that an appropriate response can be provided in a timely manner. The time instance at which the state of the process changes is referred to as the *changepoint*, and the challenge is that it is not known in advance. This is known as the *sequential (quickest) changepoint detection problem*. For lack of space, from now on we will focus only on the basic *iid* version of this problem; a general non-*iid* case is surveyed, e.g., in [16], [17].

Suppose one is able to sequentially collect a series of independent random observations,  $\{X_n\}_{n \geq 1}$ , such that  $X_1, \dots, X_\nu$

Manuscript received August 30, 2012; revised November 18, 2012; accepted December 05, 2012. Date of publication December 12, 2012; date of current version January 22, 2013. This work was supported in part by the U.S. National Science Foundation under Grant CCF-0830419 and the U.S. Defense Threat Reduction Agency under Grant HDTRA1-10-1-0086 at the University of Southern California, Department of Mathematics. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Venugopal Veeravalli.

A. G. Tartakovsky and G. Sokolov are with the Department of Mathematics and Center for Applied Mathematical Sciences, University of Southern California, Los Angeles, CA 90089-2532 USA (e-mail: tartakov@usc.edu; gsokolov@usc.edu).

A. S. Polunchenko was with the Department of Mathematics and Center for Applied Mathematical Sciences, University of Southern California, Los Angeles, CA 90089-2532 USA. He is now with the Department of Mathematical Sciences, State University of New York (SUNY) at Binghamton, Binghamton, NY 13902-6000 USA (e-mail: polunche@at.usc.edu; aleksey@binghamton.edu).

Digital Object Identifier 10.1109/JSTSP.2012.2233713

are each distributed according to a known probability density function (pdf)  $f$ , while  $X_{\nu+1}, X_{\nu+2}, \dots$  each adhere to a pdf  $g \neq f$ , also known. The time index  $\nu$  (i.e., the changepoint) is assumed unknown non-random number; for cases that regard  $\nu$  as a random variable, see, e.g., [12], [13]. One's aim is to detect that the observations' common distribution has changed. The challenge is to do so with as few observations as possible following the changepoint, subject to a tolerable limit on the risk of making a false detection.

Statistically, the problem is to sequentially differentiate between the hypotheses  $\mathcal{H}_k: \nu = k, 0 \leq k < \infty$  (i.e., that the data  $\{X_n\}_{n \geq 1}$  change their statistical profile at time instance  $\nu = k, 0 \leq k < \infty$ ) and  $\mathcal{H}_\infty: \nu = \infty$  (i.e., that no change ever occurs). To test  $\mathcal{H}_k$  against  $\mathcal{H}_\infty$  one first constructs the corresponding likelihood ratio, which for the iid scenario has the form

$$\Lambda_{k:n} = \prod_{j=k+1}^n \Lambda_j, \quad \text{where } \Lambda_j = \frac{g(X_j)}{f(X_j)},$$

and it is understood that  $\Lambda_{k:n} \equiv 1$  for  $k \geq n$ .

Next, as each new observation becomes available to test the hypotheses, the sequence  $\{\Lambda_{k:n}\}_{1 \leq k \leq n}$  is turned into a *detection statistic*. To this end, one can either use the maximum likelihood principle or the (generalized) Bayesian approach. In the former case the corresponding detection statistic is

$$V_n = \max_{1 \leq k \leq n} \Lambda_{k:n}, \quad n \geq 1, \quad (1)$$

i.e., the famous CUSUM statistic. The Bayesian statistic depends on the changepoint's prior distribution. As in our case the changepoint,  $\nu$ , is assumed unknown, the corresponding quasi-Bayesian (or generalized Bayesian) detection statistic can be defined as

$$R_n = \sum_{k=1}^n \Lambda_{k:n}, \quad n \geq 1.$$

One can view  $\{R_n\}_{n \geq 1}$  as being the average of the sequence  $\{\Lambda_{k:n}\}_{1 \leq k \leq n}$  with respect to an (improper) uniform prior distribution imposed on  $\nu$ ; see, e.g., [12], [13], [16]–[18].

Once the detection statistic is chosen, it is supplied to an appropriate sequential detection procedure. A detection procedure is a stopping time,  $T$ , which is a function of the observed data,  $\{X_n\}_{n \geq 1}$ . The meaning of  $T$  is that after observing  $X_1, \dots, X_T$  it is declared that the change is in effect. That may or may not be the case. If it is not, then  $T \leq \nu$ , and it is said that a false alarm has been sounded.

Henceforth, let  $\mathbb{P}_\nu(\cdot)$  and  $\mathbb{P}_\infty(\cdot)$  denote the probability measures, respectively, when the change occurs at time instant  $0 \leq \nu < \infty$ , and when no change ever occurs. Likewise, let  $\mathbb{E}_\nu[\cdot]$  and  $\mathbb{E}_\infty[\cdot]$  be the corresponding expectations.

Lorden [19] suggested to measure the risk of raising a false alarm via the Average Run Length (ARL) to false alarm  $\text{ARL}(T) = \mathbb{E}_\infty[T]$  and showed that the CUSUM procedure has certain asymptotically (as  $\gamma \rightarrow \inf$ ) minimax properties in the class of detection procedures

$$\Delta(\gamma) = \left\{ T: \text{ARL}(T) \geq \gamma \right\}$$

for which the ARL to false alarm is no “worse” than the desired *a priori* chosen level  $\gamma > 1$ . See also Moustakides [20] who proved that CUSUM is in fact strictly minimax with respect to Lorden's criterion for every  $\gamma > 1$ .

A practically appealing way to measure the detection speed is Pollak's [21] “worst-case” (Supremum) Average Delay to Detection (ADD), conditional on a false alarm not having been previously sounded, i.e.,

$$\text{SADD}(T) = \max_{0 \leq k < \infty} \mathbb{E}_k[T - k | T > k].$$

The minimax quickest changepoint detection problem is to find  $T_{\text{opt}} \in \Delta(\gamma)$  such that

$$\text{SADD}(T_{\text{opt}}) = \inf_{T \in \Delta(\gamma)} \text{SADD}(T) \quad \text{for all } \gamma > 1.$$

To date, this problem remains open, and only asymptotic (as  $\gamma \rightarrow \infty$ ) solutions have been obtained [21], [22].

The CUSUM chart [8] has been popular in many areas of engineering and computer science, including cybersecurity. It iteratively maximizes the log-likelihood ratio (LLR) with respect to the changepoint  $\nu$ , and stops once the maximum exceeds a certain threshold. More specifically, the CUSUM procedure is based on the statistic  $W_n = \max\{0, \log V_n\}$ , where  $V_n$  is defined in (1), which is computed recursively

$$W_n = \max\{0, W_{n-1} + \mathcal{L}_n\}, \quad n \geq 1, \quad W_0 = 0.$$

Here  $\mathcal{L}_n = \log \Lambda_n$  is the LLR. The corresponding stopping rule is

$$\mathcal{C}_h = \min\{n \geq 1: W_n \geq h\},$$

where  $h > 0$  is a detection threshold preset so as to achieve the desired level of false alarms  $\gamma > 1$ , and thus guarantee that  $\mathcal{C}_h \in \Delta(\gamma)$ . This can be achieved by setting  $h = h_\gamma \geq \log \gamma$ , since  $\text{ARL}(\mathcal{C}_h) \geq e^h$  for any  $h > 0$  [19]. For large values of  $\gamma$  more “careful” selection of  $h$  is possible [17].

Consider now a context in which it is of utmost importance to detect the change as quickly as possible, even at the expense of raising many false alarms (using a repeated application of the same stopping rule) before the change occurs. Put otherwise, in exchange for the assurance that the change will be detected with maximal speed, we agree to go through a “storm” of false alarms along the way (the false alarms are ensued from repeatedly applying the same detection rule, starting from scratch after each false alarm). This scenario is shown in Fig. 1.

Formally, let  $T_1, T_2, \dots$  be sequential independent repetitions of the stopping time  $T$ , and let  $T_j = T_1 + T_2 + \dots + T_j, j \geq 1$ , be the time of the  $j$ -th alarm. Define  $I_\nu = \min\{j \geq 1: T_j > \nu\}$ . In other words,  $T_{I_\nu}$  is the time of detection of a true change that occurs at  $\nu$  after  $I_\nu - 1$  false alarms have been raised. Write

$$\text{STADD}(T) = \lim_{\nu \rightarrow \infty} \mathbb{E}_\nu[T_{I_\nu} - \nu]$$

for the limiting value of the average delay to detection referred to as the *Stationary Average Delay to Detection* (STADD). The multi-cyclic changepoint detection problem is to find  $T_{\text{opt}} \in \Delta(\gamma)$  such that

$$\text{STADD}(T_{\text{opt}}) = \inf_{T \in \Delta(\gamma)} \text{STADD}(T) \quad \text{for every } \gamma > 1.$$

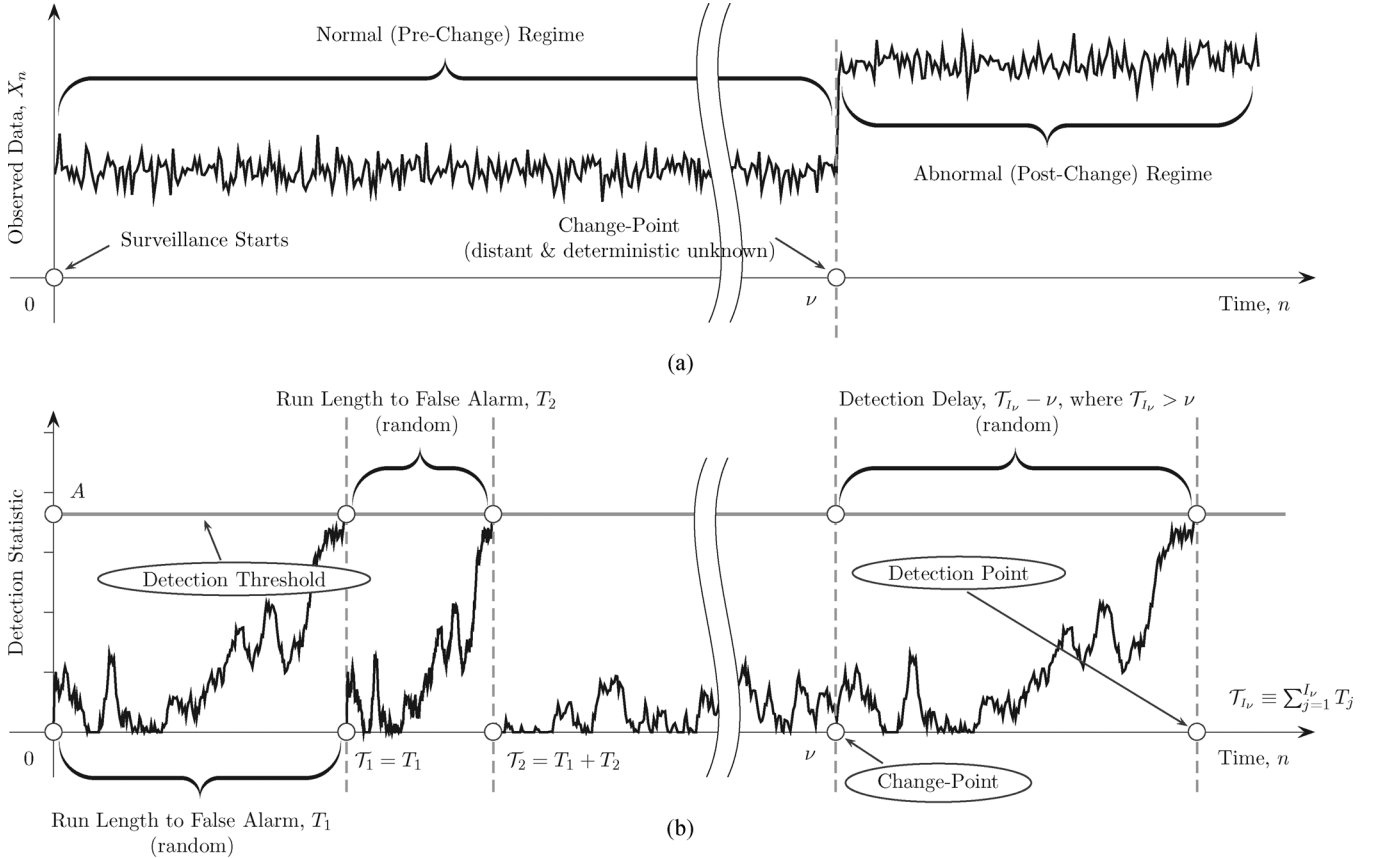


Fig. 1. Multi-cyclic changepoint detection in a stationary regime. (a) An example of the behavior of a process of interest with a change in mean at time  $\nu$ . (b) Typical behavior of the detection statistic in the multi-cyclic mode.

This formulation is instrumental in detecting a change that takes place in a distant future (i.e.,  $\nu$  is large), and is preceded by a stationary flow of false detections. Such scenarios are a commonplace in the area of computer network anomaly detection.

As has been shown by Pollak and Tartakovsky [15], the so-called Shiryaev–Roberts (SR) procedure [13], [14] is *exactly* optimal for every  $\gamma > 1$  with respect to the stationary average detection delay  $\text{STADD}(T)$ . Thus, in the multi-cyclic setting the SR procedure is a better alternative to the popular CUSUM and EWMA schemes.

The SR rule stops at time instance

$$\mathcal{S}_A = \min\{n \geq 1: R_n \geq A\},$$

where the SR statistic is given by the recursion

$$R_n = (1 + R_{n-1})\Lambda_n, \quad n \geq 1, \quad R_0 = 0.$$

Here  $A > 0$  is a detection threshold set *a priori* so as to ensure  $\mathcal{S}_A \in \Delta(\gamma)$  for a desired  $\gamma > 1$ . It can be easily shown [23] that  $\text{ARL}(\mathcal{S}_A) \geq A$  for all  $A > 0$ , so choosing the detection threshold as  $A_\gamma = \gamma$  will guarantee  $\mathcal{S}_A \in \Delta(\gamma)$ . A very accurate asymptotic approximation  $\text{ARL}(\mathcal{S}_A) \sim A/v$ ,  $A \rightarrow \infty$  is also possible, where  $0 < v < 1$  is a constant which is a subject of renewal theory. See, e.g., [23].

### III. TRANSITION TO CYBERSECURITY

The above somewhat abstract introduction to sequential changepoint detection is straightforward to put in the context

of anomaly detection in computer network traffic. As network anomalies typically take place at *unknown* points in time and entail changes in the traffic's statistical properties, it is intuitively appealing to formulate the problem of computer network anomaly detection as that of a quickest changepoint detection: to detect changes in the statistical profile of network traffic as rapidly as possible, while maintaining a tolerable level of the risk of making a false detection.

It is common that in practice neither pre- nor post-anomaly distributions are known. As a result, traffic's pre- and post-anomaly profile is poorly understood, and one can no longer rely on the likelihood ratios. Hence, an alternative approach is required. Let us first consider a typical behavior of the CUSUM and SR statistics. As long as the observed sequence  $\{X_n\}_{n \geq 1}$  is in the normal mode, the detection statistics  $\{W_n\}_{n \geq 1}$  and  $\{R_n\}_{n \geq 1}$  behave as if they were “afraid” of approaching the detection thresholds  $h$  and  $A$  respectively (although it is still possible that the thresholds will be crossed, in which case a false alarm will be raised). However, as soon as  $X_{\nu+1}$  – the first data point affected by an anomaly – is recorded, the behavior of  $W_n$  and  $R_n$  changes completely, so that they now eagerly try to hit the thresholds. Formally, this means that  $\mathbb{E}_\infty[\mathcal{L}_n] < 0$  and  $\mathbb{E}_\nu[\mathcal{L}_n] > 0$ ,  $\nu < n$ . That is, the detection statistic has a negative drift under the normal regime, and a positive drift in an anomaly situation. A typical behavior of the detection statistic in pre- and post-change regimes is shown in Fig. 2.

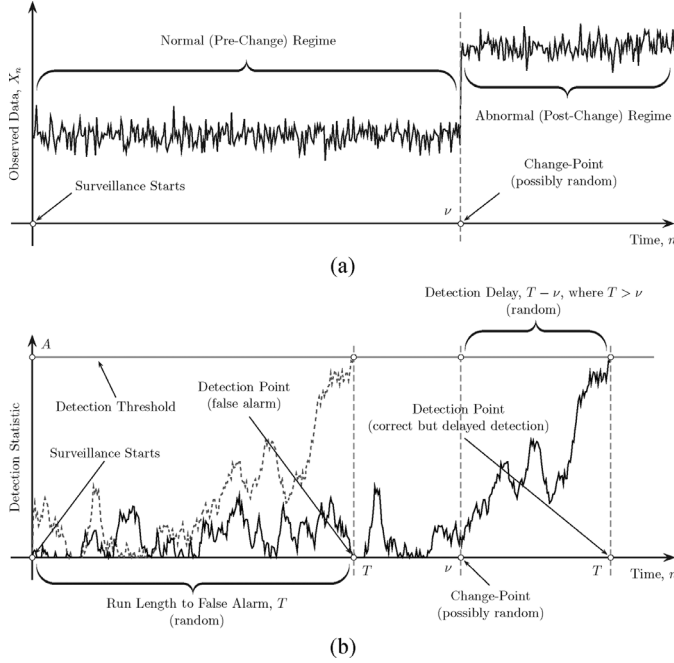


Fig. 2. Typical run of the detection statistic. (a) Observed data. (b) Two possible terminal decisions: either a false alarm (dashed), or a correct but delayed detection (solid).

Consider now the following score-based modification of the SR procedure

$$\tilde{R}_n = (1 + \tilde{R}_{n-1})e^{S_n}, \quad n \geq 1, \quad \tilde{R}_0 = 0$$

with the corresponding stopping time being

$$\tilde{S}_A = \min\{n \geq 1: \tilde{R}_n \geq A\},$$

where  $A > 0$  is an *a priori* chosen detection threshold. Similarly for CUSUM,

$$\tilde{W}_n = \max\{0, \tilde{W}_{n-1} + S_n\}, \quad n \geq 1, \quad \tilde{W}_0 = 0$$

with the corresponding stopping time being

$$\tilde{C}_h = \min\{n \geq 1: \tilde{W}_n \geq h\}, \quad h \geq 0.$$

Here  $S_n(X_1, \dots, X_n)$  are the selected score functions. Clearly, so long as

$$\mathbb{E}_\infty[S_n(X_1, \dots, X_n)] < 0 \quad \text{and} \quad \mathbb{E}_\nu[S_n(X_1, \dots, X_n)] > 0,$$

for all  $\nu \geq 0$ , the SR and CUSUM detection procedures designed using such score functions in place of the likelihood ratio will work, though they will not be optimal anymore. Their behavior will be similar to that shown in Fig. 2. Score functions  $S_n$  can be chosen in a number of ways and each particular choice depends crucially on the expected type of change. In the applications of interest, the detection problem can be usually reduced to detecting changes in mean values along with variances (mean and variance shifts).

Let

$$\begin{aligned} \mu_\infty &= \mathbb{E}_\infty[X_n], \quad \sigma_\infty^2 = \text{Var}_\infty[X_n] \\ \text{and} \quad \mu &= \mathbb{E}_0[X_n], \quad \sigma^2 = \text{Var}_0[X_n] \end{aligned}$$

denote the pre- and post-anomaly mean values and variances, respectively. Write  $Y_n = (X_n - \mu_\infty)/\sigma_\infty$  for the centered and scaled observation at time  $n$ . In the real-world applications the pre-change parameters  $\mu_\infty$  and  $\sigma_\infty^2$  are estimated from the training data and periodically re-estimated due to the non-stationarity of network traffic at large time-scales. We suggest the score  $S_n$  of the linear-quadratic form

$$S_n(Y_n) = C_1 Y_n + C_2 Y_n^2 - C_3, \quad (2)$$

where  $C_1$ ,  $C_2$  and  $C_3$  are positive design numbers assuming for concreteness that the change leads to an increase in both mean and variance. In the case where the variance either does not change or changes relatively insignificantly compared to the change in mean, the coefficient  $C_2$  may be set to zero. In the opposite case where the mean changes only slightly compared to the variance, we take  $C_1 = 0$ . The first case appears to be typical for many cybersecurity applications, for example for ICMP and UDP Denial-of-Service (DoS) attacks (see [4], [5] where the linear score-based CUSUM has been proposed). However, in certain cases, such as the one considered below in Section IV, both the mean and variance change quite significantly.

Note that the score given by (2) with

$$C_1 = \delta q^2, \quad C_2 = \frac{(1 - q^2)}{2}, \quad C_3 = \frac{\delta^2 q^2}{2} - \log q, \quad (3)$$

where  $q = \sigma_\infty/\sigma$ ,  $\delta = (\mu - \mu_\infty)/\sigma_\infty$ , is optimal if pre- and post-change distributions are Gaussian with known putative values  $\mu$  and  $\sigma^2$ . This is true because in the latter case  $S_n$  is the log-likelihood ratio. If one believes in the Gaussian model (which sometimes is the case), then selecting  $q = q_0$  and  $\delta = \delta_0$  with some design values  $q_0$  and  $\delta_0$  provides reasonable operating characteristics for  $q < q_0$  and  $\delta > \delta_0$  and optimal characteristics for  $q = q_0$  and  $\delta = \delta_0$ . However, it is important to emphasize that the proposed score-based SR procedure does not assume that the observations have Gaussian pre- and post-change distributions.

Further improvement may be achieved by using either mixtures or adaptive versions with generalized likelihood ratio-type statistics [19], [23].

Based on the previous reasoning (see Section II) we expect the multi-cyclic score-based SR procedure to perform better than the analogous CUSUM procedure.

#### IV. A CASE STUDY

We now present the results of testing the proposed detection algorithms on a real Distributed DoS (DDoS) attack, namely, SYN flood attack. The aim of this attack is to congest the victim's link with a series of SYN requests so as to have the victim's machine exhaust all of its resources and stop responding to legitimate traffic. This kind of an attack clearly creates a volume-type anomaly in the victim's traffic flow. The data is courtesy of the Los Angeles Network Data Exchange and Repository (LANDER) project (see <http://www.isi.edu/ant/lander>). Specifically, the trace is flow data captured by Merit Network Inc. (see <http://www.merit.edu>). The attack is on a University of Michigan IRC server. It starts at roughly 550 seconds into the trace and has a duration of 10 minutes. The attacked IP is

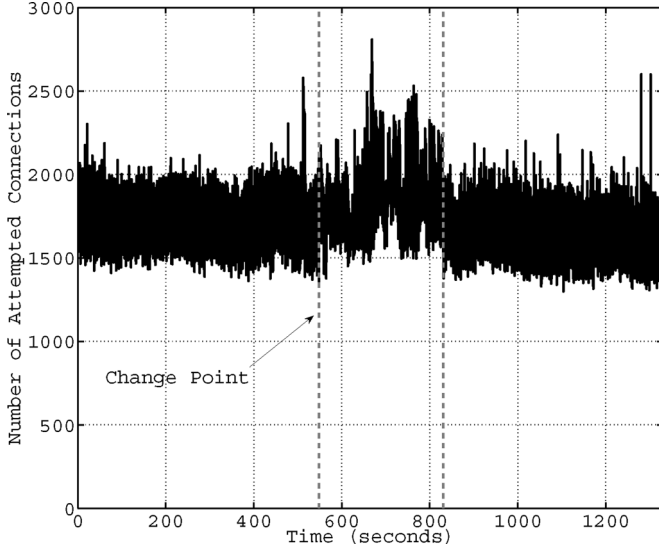
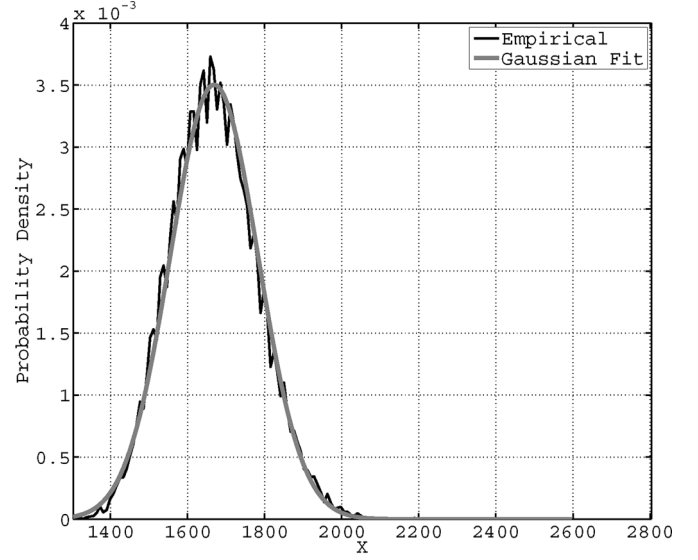


Fig. 3. SYN flood attack: number of attempted connections.

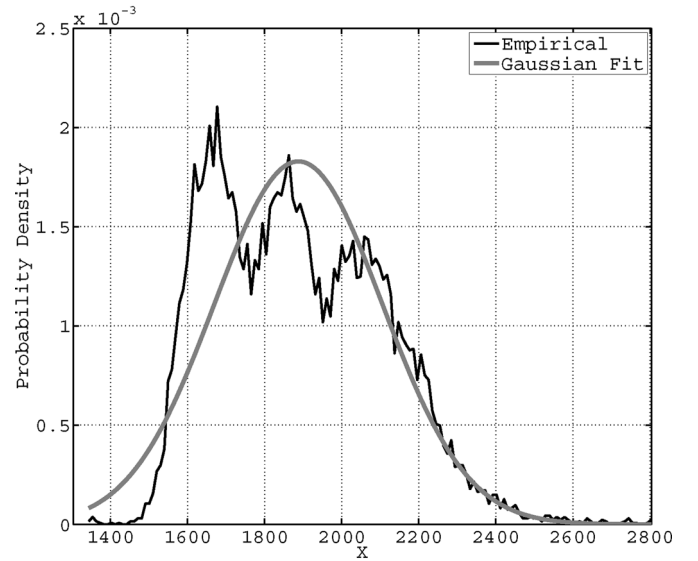
anonymized to 141.213.238.0. Fig. 3 shows the number of attempted connections or the connections birth rate as a function of time. While the attack can be seen to the naked eye, it is not completely clear when it starts. In fact, there is a spike in the data (fluctuation) before the attack. Also, controlling the false alarm rate with an automatic detection system is a challenge.

We used the number of connections during 20 msec batches as the observations  $X_n$ . We estimated the connections birth rate average and variance for legitimate traffic and for attack traffic; in both cases, to estimate the average we used the usual sample mean, and to estimate the variance we used the usual sample variance. For legitimate traffic, the average is about  $\mu_\infty = 1669.09$  connections per 20 msec, and the standard deviation is in the neighborhood of  $\sigma_\infty = 113.884$  connections per 20 msec. For attack traffic, the numbers are  $\mu = 1887.56$  and  $\sigma = 218.107$ , respectively. We can now see the effect of the attack: it leads to a considerable increase in the mean and standard deviation of the connections birth rate.

We now perform a basic statistical analysis of the connections birth rate distribution. Fig. 4 shows the empirical densities of the connections birth rate for legitimate and attack traffic. It so happens that for given data, legitimate traffic appears to resemble the Gaussian process. However, for attack traffic, the distribution is not as close to Gaussian. We have implemented the score-based multi-cyclic SR and CUSUM procedures with the linear-quadratic score (2). When choosing the design parameters we assume the Gaussian model for pre-attack traffic, which agrees with the conclusions drawn above following the basic statistical analysis of the data. Thus, parameters  $C_1$ ,  $C_2$ , and  $C_3$  are chosen according to formulas (3) with  $q_0 = q \approx 0.52$  and to allow for detection of fainter attacks  $\delta_0 \approx 1.5$  (versus the estimated attack value  $\delta \approx 1.9$ ). We set the detection thresholds  $A \approx 1.9 \times 10^3$  and  $h \approx 6.68$  so as to ensure the same level of ARL at approximately 500 samples (i.e., 10 sec) for both procedures. The thresholds are estimated using Monte Carlo simulations assuming the empirical pre-change distribution learned from the data. Specifically, we took  $10^5$  samples from the empirical pre-change distribution and simulated the behavior of the



(a)



(b)

Fig. 4. SYN flood attack: connections birth rate pdf with a Gaussian fit. (a) Legitimate (pre-attack) traffic. (b) Attack traffic.

respective detection statistics and procedures while adjusting the thresholds until observing the desired ARL.

The detection process is illustrated in Figs. 5 and 6. Fig. 5 shows a relatively long run (taking into account the sampling rate 20 msec) of the SR statistic with several false alarms and then the true detection of the attack with a very small detection delay (at the expense of raising many false alarms along the way). Recall that the whole idea of this paper is to set the detection thresholds low enough in order to detect attacks very quickly with minimal delays, which unavoidably leads to multiple false alarms prior to the attack starts. These false alarms should be filtered by a specially designed algorithm, as has been suggested in [15] and will be further discussed in Section V.

Fig. 6(a) shows the behavior of the logarithm of the SR statistic shortly prior to the attack and right after the attack starts till its detection, which happens when the statistic crosses the threshold. Fig. 6(b) shows the same for the CUSUM

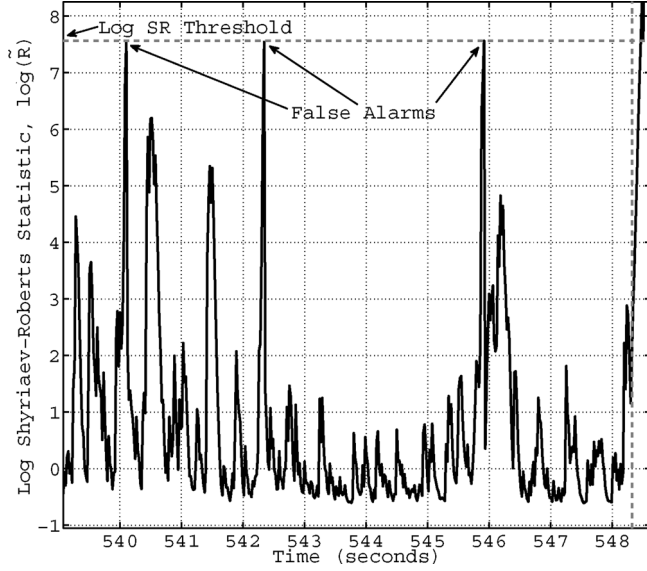


Fig. 5. SYN flood attack: long run of the Shyriaev–Roberts procedure; logarithm of the SR statistic versus time.

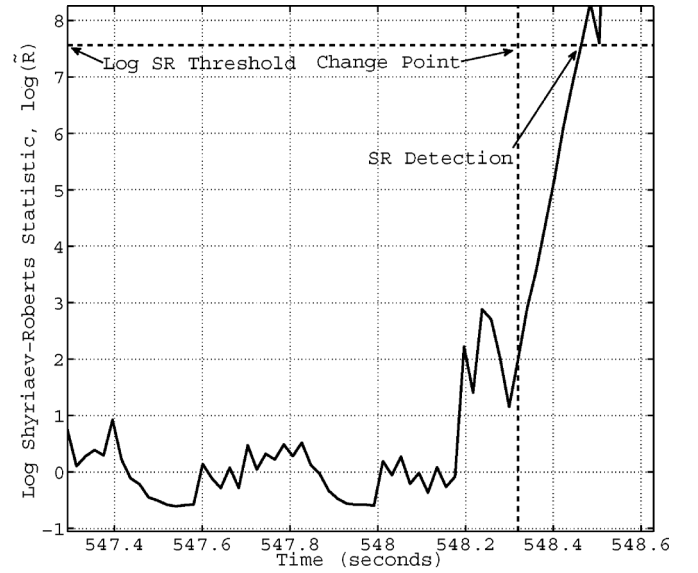
statistic. We see that both procedures successfully detect the attack with very small delays, though at the expense of raising false alarms along the way, as shown in Fig. 5 and discussed above. For both procedures we observed approximately 7 false alarms per 1000 samples. The detection delay for the repeated SR procedure is roughly 0.14 seconds (or 7 samples), and for the CUSUM procedure the delay is about 0.21 seconds (or 10 samples). Thus, the SR procedure is better, as expected.

## V. FURTHER DISCUSSION

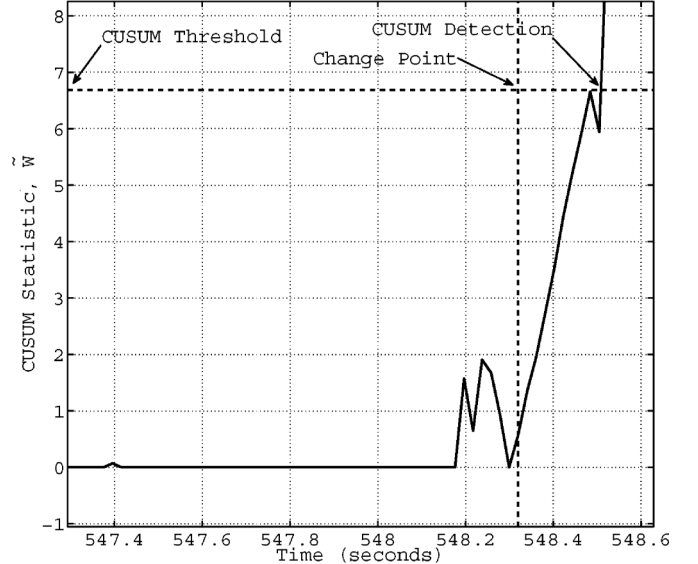
Since in real life legitimate traffic dominates, the idea of comparing various anomaly-based IDS-s using the multi-cyclic approach and the stationary average detection delay is a natural fit for cybersecurity applications. However, it is worthwhile to remark on a possible way to enhance the potential of changepoint detection techniques as applied to cybersecurity. Any changepoint detection method is subject to the following drawback: instantaneous detection is not an option, unless the false alarm risk is high. Hence, though changepoint detection schemes are computationally inexpensive, in practice, employing one such scheme alone may not be a good idea, since it will be overflowed with false alarms. The simplest solution is to increase detection thresholds dramatically, but this will lead to an increase of the detection delay.

Here comes an interesting opportunity: What if one could combine changepoint detection techniques with others that offer very low false alarm rate, but are too heavy to use at line speeds? Do such synergistic anomaly detection systems exist, and how can they be integrated?

As an answer, consider complementing a changepoint detection-based anomaly detector with a flow-based signature IDS that examines the traffic’s spectral profile. For an example of such signature-flow-based method, see, e.g., [24]–[27]. The principal idea is to employ the Fourier transform to obtain the corresponding spectral characteristics of the passing traffic. This idea can be used in conjunction with the changepoint



(a)



(b)

Fig. 6. Detection of the SYN flood attack by the SR and CUSUM procedures. (a) By the SR procedure. (b) By the CUSUM procedure.

detection-based anomaly detector for both rejection of false alarms and confirmation of true detections. Higher computational complexity of the spectral-signature based detector is compensated by the preliminary changepoint anomaly based algorithm; the latter triggers the former only when there is a suspicion of an anomaly may be taking place in the network link of interest. For practical purposes the mean time between false alarms of the changepoint based anomaly IDS can be taken as small as a few seconds, as it was in the experiments presented in the previous section. We believe that such an alliance of the changepoint anomaly- and the spectral-signature-based detectors can significantly improve the whole system’s overall performance reducing the false alarm rate to the minimum and at the same time guaranteeing very small detection delays.

## VI. CONCLUSION

We addressed the problem of rapid anomaly detection in computer network traffic. Approaching the problem statistically, namely, as that of sequential changepoint detection, we proposed a new anomaly detection method. The method is based on the multi-cyclic (repeated) Shiryaev–Roberts detection procedure where the likelihood ratio is replaced with the linear-quadratic score. This is done because in real-world network security applications both pre-attack and post-attack distributions are different from hypothesized distributions such as Gaussian or Poisson. Like many changepoint detection schemes, our method is also of practically no computational complexity and easy to implement. However, what distinguishes the SR procedure is its exact multi-cyclic optimality in a simple change detection problem where densities are known, a property that such techniques as the SPRT, the CUSUM chart, or the EWMA scheme lack. Hence, one may conjecture that the score-based SR detection algorithm is a better cyber “watchdog”. To support this conjecture, we conducted a case study using a real SYN flood attack. The score-based multi-cyclic SR algorithm outperformed the multi-cyclic CUSUM procedure. Lastly, as a possible improvement of any changepoint detection-based anomaly detector, we proposed to complement the latter with a signature-based spectral IDS. This approach will allow to filter false alarms reducing the false alarm rate to a minimum and simultaneously guaranteeing prompt detection of real attacks.

## ACKNOWLEDGMENT

The authors would like to thank Dr. C. Papadopoulos (Colorado State University, Department of Computer Science) and Dr. J. Heidemann (University of Southern California, Information Sciences Institute) for help with obtaining real data. The authors are also grateful to two anonymous referees whose comments helped to improve the paper.

## REFERENCES

- [1] J. Ellis and T. Speed, *The Internet Security Guidebook: From Planning to Deployment*. New York: Academic, 2001.
- [2] G. Thattai, U. Mitra, and J. Heidemann, “Parametric methods for anomaly detection in aggregate traffic,” *IEEE/ACM Trans. Networking*, vol. 19, no. 2, pp. 512–525, Apr. 2011.
- [3] G. Thattai, U. Mitra, and J. Heidemann, “Detection of low-rate attacks in computer networks,” in *Proc. of the 11th IEEE Global Internet Symp.*, Phoenix, AZ, Apr. 2008, pp. 1–6.
- [4] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blažek, and H. Kim, “A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods,” *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3372–3382, Sep. 2006.
- [5] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blažek, and H. Kim, “Detection of intrusions in information systems by sequential changepoint methods (with discussion),” *Statistical Methodol.*, vol. 3, no. 3, pp. 252–340, 2006.
- [6] A. G. Tartakovsky, B. L. Rozovskii, and K. Shah, “A nonparametric multichart CUSUM test for rapid intrusion detection,” in *Proc. Joint Statist. Meetings*, Minneapolis, MN, Aug. 2005.
- [7] A. Wald, *Sequential Analysis*. New York: Wiley, 1947.
- [8] E. S. Page, “Continuous inspection schemes,” *Biometrika*, vol. 41, no. 1, pp. 100–115, 1954.

- [9] S. Roberts, “Control chart tests based on geometric moving averages,” *Technometrics*, vol. 1, no. 3, pp. 239–250, Aug. 1959.
- [10] H. V. Poor and O. Hadjiladis, *Quickest Detection*. Cambridge, U.K.: Cambridge Univ. Press, 2008.
- [11] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [12] A. N. Shiryaev, “The problem of the most rapid detection of a disturbance in a stationary process,” *Soviet Math. Dokl.*, vol. 2, pp. 795–799, 1961.
- [13] A. N. Shiryaev, “On optimum methods in quickest detection problems,” *Theory of Probability and its Applicat.*, vol. 8, no. 1, pp. 22–46, Jan. 1963.
- [14] S. Roberts, “A comparison of some control chart procedures,” *Technometrics*, vol. 8, no. 3, pp. 411–430, Aug. 1966.
- [15] M. Pollak and A. G. Tartakovsky, “Optimality properties of the Shiryaev–Roberts procedure,” *Statistica Sinica*, vol. 19, no. 4, pp. 1729–1739, 2009.
- [16] A. G. Tartakovsky and G. V. Moustakides, “State-of-the-art in Bayesian changepoint detection,” *Sequential Anal.*, vol. 29, no. 2, pp. 125–145, Apr. 2010.
- [17] A. S. Polunchenko and A. G. Tartakovsky, “State-of-the-art in sequential change-point detection,” *Methodol. and Comput. in Appl. Probability*, vol. 14, no. 3, pp. 649–684, 2012.
- [18] M. A. Girschick and H. Rubin, “A Bayes approach to a quality control model,” *Ann. Math. Statist.*, vol. 23, no. 1, pp. 114–125, 1952.
- [19] G. Lorden, “Procedures for reacting to a change in distribution,” *Ann. Math. Statist.*, vol. 42, no. 6, pp. 1897–1908, 1971.
- [20] G. V. Moustakides, “Optimal stopping times for detecting changes in distributions,” *Ann. Statist.*, vol. 14, no. 4, pp. 1379–1387, 1986.
- [21] M. Pollak, “Optimal detection of a change in distribution,” *Ann. Statist.*, vol. 13, no. 1, pp. 206–227, 1985.
- [22] A. G. Tartakovsky, M. Pollak, and A. S. Polunchenko, “Third-order asymptotic optimality of the generalized Shiryaev–Roberts changepoint detection procedures,” *Theory of Probab. and its Applicat.*, vol. 56, no. 3, pp. 457–484, Aug. 2012.
- [23] M. Pollak, “Average run lengths of an optimal method of detecting a change in distribution,” *Ann. Statist.*, vol. 15, no. 2, pp. 749–779, 1987.
- [24] A. Hussain, J. Heidemann, and C. Papadopoulos, “A framework for classifying denial of service attacks,” in *Proc. Conf. Applicat., Technol., Architectures, Protocols for Comput. Commun.*, Karlsruhe, Germany, Aug. 2003, pp. 99–110.
- [25] X. He, C. Papadopoulos, J. Heidemann, U. Mitra, U. Riaz, and A. Hussain, “Spectral analysis of bottleneck traffic,” Univ. of Southern California, Dept. of Comput. Sci., 2005, Tech. Rep. USC-CSD-TR-05-854.
- [26] A. Hussain, J. Heidemann, and C. Papadopoulos, “Identification of repeated denial of service attacks,” in *Proc. 25th IEEE Int. Conf. Comput. Commun.*, Barcelona, Spain, Apr. 2006, pp. 1–15.
- [27] X. He, C. Papadopoulos, J. Heidemann, U. Mitra, and U. Riaz, “Remote detection of bottleneck links using spectral and statistical methods,” *Comput. Netw.*, vol. 53, no. 3, pp. 279–298, Feb. 2009.



**Alexander G. Tartakovsky** (SM’02) is Professor of Mathematics in the Department of Mathematics and the Associate Director of the Center for Applied Mathematical Sciences at the University of Southern California, Los Angeles. He is also Vice President of Argo Science Corp., Rolling Hills Estates, CA. His research interests include theoretical and applied statistics; applied probability; sequential analysis; changepoint detection phenomena; adaptive, minimax and robust methods for overcoming prior uncertainty; statistical image and signal processing;

visual tracking; detection and tracking of objects in radar and infrared search and track systems; information integration/fusion; cybersecurity (rapid intrusion detection); and detection and tracking of malicious activity. He is the author of one book (“Sequential Methods in the Theory of Information Systems”) and over 100 articles in the areas indicated above. Dr. Tartakovsky obtained a Ph.D. in statistics and information theory from Moscow Institute of Physics and Technology (Russia) in 1981, and an advanced Doctor-of-Science degree in statistics and control from Moscow Institute of Physics and Technology (Russia) in 1990. He is a Fellow of the Institute of Mathematical Statistics, a senior member of IEEE and a member of many scientific societies such as SIAM, INFORMS, Bernoulli, ISI, etc. Dr. Tartakovsky is a recipient of the 2007 Abraham Wald Award in Sequential Analysis.



**Aleksey S. Polunchenko** is currently Assistant Professor in the Department of Mathematical Sciences at the State University of New York (SUNY) at Binghamton, Binghamton, New York, USA. His research spans theoretical and applied statistics, applied probability, sequential analysis, quickest changepoint detection, mathematical finance, numerical methods, computer vision, cybersecurity, and parallel computing. Dr. Polunchenko obtained a Ph.D. in Applied Mathematics and an M.Sc. in Mathematical Finance from the University of Southern California (USC), Los Angeles, California, USA in 2009, followed by a three-year postdoctoral research fellowship at USC in the group of Dr. Tartakovsky. Dr. Polunchenko also received an M.Sc. in Applied Mathematics and Physics from Moscow Institute of Physics and Technology, Moscow, Russia in 2004. Dr. Polunchenko is a recipient of the Institute of Mathematical Statistics' 2009 *Laha Travel Award*. He also holds various professional certifications from Microsoft and SAS.



**Grigory Sokolov** is a graduate student in the Department of Mathematics at the University of Southern California, Los Angeles, California, USA. His research interests include theoretical and applied statistics, applied probability, sequential analysis, quickest changepoint detection, numerical analysis, cybersecurity, and parallel programming. Sokolov received a B.S. in applied mathematics and physics in 2006, followed by a M.Sc. in applied mathematics and physics with specialization in information processing and control systems from Moscow Institute of Physics and Technology, Moscow, Russia in 2008.