Available online at www.sciencedirect.com
SCIENCE DIRECT
Statistical Methodology 3 (2006) 315–319
Statistical Methodology
Official Journal of the International Indian Statistical Association
www.elsevier.com/locate/stamet

Discussion

A discussion on 'Detection of intrusions in information systems by sequential change-point methods' by Tartakovsky, Rozovskii, Blažek, and Kim

Seong-Hee Kim[a], James R. Wilson[b,*]

[a] *School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA*
[b] *Department of Industrial Engineering, North Carolina State University, Raleigh, NC, 27695-7906, USA*

Received 29 June 2005; accepted 29 June 2005

Abstract

The discussion focuses on issues arising in practical applications of the CUSUM procedures developed by Tartakovsky et al. to detect changes in multichannel sensor systems. These issues are illustrated with a data set from the MIT Lincoln Laboratory that is used to estimate the parameters of procedures for detecting denial-of-service attacks.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Change-point detection; CUSUM procedures; Sequential tests; Intrusion detection; Denial of service

1. Introduction

The paper by Tartakovsky et al. [7] is a welcome contribution to the literature on sequential procedures for change-point detection; and the potential applications of the methodology developed in this paper should extend well beyond the area of intrusion detection in information systems. In this note, the discussion is focused on issues arising when the proposed multichannel CUSUM procedures are implemented in practical applications—especially in circumstances that require rapid calibration of the procedures and that do not allow extensive preliminary experimentation on training data sets to establish parameter values for the procedures yielding

DOI of original article: 10.1016/j.stamet.2005.05.003.
* Corresponding address: Department of Industrial Engineering, North Carolina State University, Campus Box 7906, Raleigh, NC 47695-7906, USA. Tel.: +1 919 515 6415; fax: +1 919 515 5281.
E-mail addresses: skim@isye.gatech.edu (S.-H. Kim), jwilson@ncsu.edu (J.R. Wilson).

1572-3127/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.stamet.2005.06.006

a prespecified false alarm rate or to estimate the average detection delay under appropriate hypotheses on system status.

As pointed out in [7], the LR-CUSUM procedure is rarely applicable in practice because users do not often know the exact form of the joint distribution of the basic random variables $\{X_i(j) : j = 1, 2, \ldots\}$ observed in channel i for $i = 1, \ldots, N$. In particular, users generally lack detailed knowledge of: (a) the marginal distribution(s) of the process $\{X_i(j) : j = 1, 2, \ldots\}$; and (b) the stochastic dependencies among successive observations of this process. Therefore, we focus on practical applications of the NP-CUSUM and B-CUSUM algorithms.

The NP-CUSUM algorithm requires estimation of the parameters μ_i, c_i, and w_i in channel i for $i = 1, \ldots, N$. Unlike typical statistical process control (SPC) schemes that require estimating some parameters from an in-control process only, the NP-CUSUM procedure also requires estimating some parameters from an out-of-control process—specifically, ε together with θ_i or δ_i for $i = 1, \ldots, N$. As enumerated below, there are several problems inherent in this situation.

1. When an attack is obvious, typically it is difficult to obtain a training data set corresponding to an out-of-control (intrusion) status such that the sample size is large enough to yield sufficiently accurate estimates of the required parameters.
2. On the other hand if the attack is subtle, then it will be hard to identify which parts of the training data set(s) correspond to in-control and out-of-control conditions.
3. More importantly, the size of the shift in the mean of a monitored process can range from small to large values, and it is not guaranteed that the size of the shift will be the same for different instances of the same type of intrusion. As a consequence, c_i will vary depending on the sampled out-of-control data set. Therefore, it seems more reasonable to "tune" the NP-CUSUM algorithm so that it is most sensitive to a specific shift amount. This approach is used to set the parameter K in the Tabular CUSUM procedure for i.i.d. normal data [4]; and in the Tabular CUSUM procedure, K is analogous to the parameter c_i in the NP-CUSUM algorithm [7].
4. Another problem is that we still need to determine the threshold h by trial and error. Many SPC charts for autocorrelated or multivariate data that have been proposed in the recent literature determine their control limits by simulation based on a specific probabilistic model for the in-control process. However, if the assumptions underlying that model are violated, then the charts will not work as advertised; and thus there have been efforts to develop model-free SPC procedures [3,6]. Sometimes it is difficult to represent data in intrusion detection using a specific probabilistic model, as illustrated by the example given below; and, more seriously, determining the threshold h by trial and error cannot be done automatically in real time as pointed out by Tartakovsky et al. [7].

In the case of the B-CUSUM algorithm, issues 1–4 also apply to the CUSUM threshold h_b and the quantization threshold t_i for $i = 1, \ldots, N$.

In [7], the performance of the NP-CUSUM and B-CUSUM algorithms is demonstrated for Poisson-distributed data and for data from the MIT Lincoln Laboratory representing traffic in a real network that is subjected to a TCP SYN flooding attack. The latter data set seems to be relatively free of anomalous behavior that might complicate the application of the CUSUM procedures developed in [7]. However, some traffic data related to other types of network intrusion exhibit much less tractable characteristics as detailed below.

The MIT Lincoln Laboratory simulated the environment of a real network to provide a test-bed of data sets for comprehensive evaluation of the performance of various intrusion detection systems. Ye et al. [9,10] and Park [5] derive event intensity (arrival-rate) data from log files generated by the Basic Security Module (BSM) of a Sun SPARC 10 workstation running the

Download English Version:

https://daneshyari.com/en/article/1151231

Download Persian Version:

https://daneshyari.com/article/1151231

Daneshyari.com