REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
11-07-2014	Book	-

4. TITLE AND SUBTITLE
Sequential Analysis: Hypothesis Testing and Changepoint Detection

5a. CONTRACT NUMBER
W911NF-14-1-0246

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
Alexander Tartakovsky, Igor Nikiforov, Michele Basseville

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
University of Connecticut - Storrs
438 Whitney Road Ext., Unit 1133

Storrs, CT 06269 -1133

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
65367-MA.1

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
The main focus of this book is on a systematic development of the theory of sequential hypothesis testing (Part I) and changepoint detection (Part II). In Part III, we briefly describe certain important applications where theoretical results can be used efficiently, perhaps with some reasonable modifications. We review recent accomplishments in hypothesis testing and changepoint detection both in decision-theoretic (Bayesian) and non-decision-theoretic (non-Bayesian) contexts. The emphasis is not only on more traditional binary hypotheses but also on substantially more difficult multiple decision problems. Scenarios with simple hypotheses and more realistic cases of (two and finitely

15. SUBJECT TERMS
Sequential hypothesis testing, quickest changepoint detection

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Alexander Tartakovsky
UU	UU	UU	UU		19b. TELEPHONE NUMBER: 860-486-5804

Report Title

Sequential Analysis: Hypothesis Testing and Changepoint Detection

ABSTRACT

The main focus of this book is on a systematic development of the theory of sequential hypothesis testing (Part I) and changepoint detection (Part II). In Part III, we briefly describe certain important applications where theoretical results can be used efficiently, perhaps with some reasonable modifications. We review recent accomplishments in hypothesis testing and changepoint detection both in decision-theoretic (Bayesian) and non-decision-theoretic (non-Bayesian) contexts. The emphasis is not only on more traditional binary hypotheses but also on substantially more difficult multiple decision problems. Scenarios with simple hypotheses and more realistic cases of (two and finitely many) composite hypotheses are considered and treated in detail. While our major attention is on more practical discrete-time models, since we strongly believe that “life is discrete in nature” (not only due to measurements obtained from devices and sensors with discrete sample rates), certain continuous-timemodels are also considered once in a while, especially when general results can be obtained very similarly in both cases. It should be noted that although we have tried to provide rigorous proofs of the most important results, in some cases we included heuristic argument instead of the real proofs as well as gave references to the sources where the proofs can be found.

Sequential Analysis: Hypothesis Testing and Changepoint Detection

Alexander G. Tartakovsky
Department of Statistics
University of Connecticut
Storrs, CT USA

&

Igor V. Nikiforov
Université de Technologie de Troyes
Troyes, France

&

Michèle Basseville
CNRS & IRISA
Rennes, France

Contents

Preface

About seventy years ago Abraham Wald, while treating the problem of testing two simple hypotheses, showed how the fixed sample size likelihood ratio test of Neyman and Pearson can be modified into the more efficient sequential scheme when observations are collected one at a time and processed on-line. This has led to the modern theory of sequential analysis developed due to a practical demand for more efficient sampling policies and summarized by A. Wald in his monograph *Sequential Analysis* published in 1947.

A separate important branch of sequential analysis is on-line surveillance, the so-called changepoint detection, the goal of which is to detect a change in distribution or anomaly quickly. More specifically, sequential changepoint detection (or quickest change/"disorder" detection) is concerned with the design and analysis of techniques for on-line detection of a change in the state of a phenomenon, subject to a tolerable limit on the risk of false alarms. An observed process of interest may unexpectedly undergo an abrupt change-of-state from "normal" to "abnormal" (or anomalous), each defined as deemed appropriate given the physical context. The sequential setting assumes the observations are made successively, and, as long as their behavior suggests that the process is in the normal state, the process is allowed to continue. However, if the state is believed to have altered, one's aim is to detect the change "as soon as possible," so that an appropriate response can be provided in a timely manner.

Historically, the subject of changepoint detection first began to emerge in the 1920s motivated by considerations of industrial quality control due to the work of Walter Shewhart who successfully brought together the disciplines of statistics, engineering, and economics and became the father of modern statistical quality control. Shewhart's work (in particular Shewhart control charts) was highlighted in his books *Economic Control of Quality of Manufactured Product* (1931) [411] and *Statistical Method from the Viewpoint of Quality Control* (1939) [412], for which he gained recognition in the statistical community, but efficient (optimal and quasi-optimal) sequential detection procedures were developed much later in the 1950–1960s after the emergence of Wald's book *Sequential Analysis* (1947) [494]. The ideas set in motion by Shewhart and Wald have formed a platform for extensive research on both theory and practice of sequential changepoint detection, starting with the seminal paper by Page (1954) where the now famous Cumulative Sum (CUSUM) detection procedure was first proposed, and followed by the series of works of Shiryaev (1961–1969) [414, 413, 415, 416, 417, 418, 419] and Lorden (1971) [271] where the first optimality results in Bayesian and non-Bayesian contexts were established.

During the past 20 years, general stochastic models appropriate for many interesting applications have been treated extensively, as theoretical foundation for asymptotic studies of properties of known sequential tests such as Wald's Sequential Probability Ratio Test (SPRT), matrix versions of this test suitable for multiple decision problems, CUSUM and Shiryaev–Roberts change detection procedures, which are known to be optimal or nearly optimal for the models with independent and identically distributed (iid) observations. Asymptotic optimality of these rules has been established under various conditions, including conventional iid and general non-iid scenarios. Novel procedures have also been proposed and studied. Multihypothesis and multichannel change detection–classification (or detection–isolation) rules have been developed and their asymptotic optimality properties have been established for iid and general non-iid models. Even for relatively simple iid models new results have been obtained, in particular toward very precise analysis via solving integral equations numerically and asymptotic analysis using renewal-theoretic and nonlinear renewal-

theoretic approaches. These numerical and asymptotic approaches are in fact complementary, since numerical solutions become very time-consuming when dealing with small error probabilities or low false alarm rates, while asymptotic approximations are usually not too accurate for high and moderate false alarm rates.

The main focus of this book is on a systematic development of the theory of sequential hypothesis testing (Part I) and changepoint detection (Part II). In Part III, we briefly describe certain important applications where theoretical results can be used efficiently, perhaps with some reasonable modifications. We review recent accomplishments in hypothesis testing and changepoint detection both in decision-theoretic (Bayesian) and non-decision-theoretic (non-Bayesian) contexts. The emphasis is not only on more traditional binary hypotheses but also on substantially more difficult multiple decision problems. Scenarios with simple hypotheses and more realistic cases of (two and finitely many) composite hypotheses are considered and treated in detail. While our major attention is on more practical discrete-time models, since we strongly believe that "life is discrete in nature" (not only due to measurements obtained from devices and sensors with discrete sample rates), certain continuous-time models are also considered once in a while, especially when general results can be obtained very similarly in both cases. It should be noted that although we have tried to provide rigorous proofs of the most important results, in some cases we included heuristic argument instead of the real proofs as well as gave references to the sources where the proofs can be found.

While there are many other interesting topics in sequential analysis such as point and interval estimation, selection/ranking, and sequential games, these important topics are out of the scope of our book. A detailed treatment of these additional sequential methods can be found, e.g., in [56, 163, 259, 312, 452].

We would like to thank many colleagues who have directly and indirectly contributed to this project. Several students and postdoctoral fellows at the University of Southern California worked on some of the problems considered in the book at different stages. Aleksey Polunchenko contributed to certain theoretical aspects and numerical methods related to the very precise analysis of minimax changepoint detection procedures as well as helped with simulations and processing real and semi-real data for computer network security applications. The joint work with Georgios Fellouris on minimax tests for discrete composite hypotheses became the basis for the corresponding sections in Part I. Greg Sokolov performed useful numerical analysis and Monte Carlo simulations of multichannel change detection procedures. Collaboration with George Moustakides, Moshe Pollak, and Venugopal Veeravalli as well as frequent discussions with them were extremely fruitful. The joint work with Lionel Fillatre on FSS multiple hypothesis testing has been used for writing Subsections 2.9.6 and 10.2.2.

Alexander Tartakovsky is thankful to various U.S. agencies (Department of Defense, Department of Energy, National Science Foundation) for supporting his work under multiple contracts.[1]

Alexander Tartakovsky wants to thank his wife, Marina Blanco, for her patience, help, and inspiration.

Igor Nikiforov is thankful to the University of Technology of Troyes for supporting this work and for the environment in which the book has been written. A preliminary version of some material of the book has been used for Master and PhD courses at the University of Technology of Troyes. The work reported in Section 11.1 has been partly supported by the SERCEL (Sociètè d'Etudes, Recherches et Constructions Électroniques), by the SAGEM (Société d'Applications Générales d'Électricité et de Mécanique, of the SAFRAN group), by the LRBA (Laboratoire de Recherches Balistiques et Aérodynamiques) and by the DGAC/DTI (Direction de la Technique et de l'Innovation, formerly known as STNA).

[1] In particular, the work of Alexander Tartakovsky was partially supported by the U.S. Air Force Office of Scientific Research under MURI grant FA9550-10-1-0569, by the U.S. Defense Threat Reduction Agency under grant HDTRA1-10-1-0086, by the U.S. Defense Advanced Research Projects Agency under grant W911NF-12-1-0034, the U.S. Army Research Office under MURI grant W911NF-06-1-0044 and under grants W911NF-13-1-0073 and W911NF-14-1-0246, and by the U.S. National Science Foundation under grants CCF-0830419, EFRI-1025043, and DMS-1221888 at the University of Southern California, Department of Mathematics and at the University of Connecticut, Department of Statistics.

Igor Nikiforov wants to thank his wife, Tatiana, for her support, understanding, and encouragement during the writing of this book.

Michèle Basseville is thankful to the Centre National de la Recherche Scientifique (CNRS) for its support and to the Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA) for the environment in which the book has been written. The work reported in Section 11.2 has been partly supported by the Eurêka projects no 1562 SINOPSYS, no 2419 FliTE and no 3341 FliTE2.

Finally, we are grateful to two anonymous referees whose comments have improved the presentation.

Los Angeles, California and Storrs, Connecticut USA[2] Alexander Tartakovsky
Troyes, France Igor Nikiforov
Rennes, France Michèle Basseville

[2]The book was completed after Alexander Tartakovsky joined the Department of Statistics, University of Connecticut at Storrs (September 2013) but most of the book was written while he was with the Department of Mathematics, University of Southern California, Los Angeles.

List of Figures

List of Tables

Notation and Symbols

Notation	Meaning
$X_t \xrightarrow[t\to\infty]{\mathsf{P}-\text{a.s.}} Y$	Almost sure convergence under P (or with probability 1).
R_n	*A posteriori* risk (APR); also minimum *a posteriori* risk (MAPR).
R_n^{st}	APR associated with stopping.
$\tilde{R}_n$	APR associated with continuation of observations.
$\rho(\delta)$	Average (or integrated) risk (AR).
ADD	Average delay to detection (detection delay).
$\mathbf{C}(\alpha)$	Class of tests with significance level α.
$[a,b]$	Closed interval.
$X_t \xrightarrow[t\to\infty]{\text{completely}} Y$	Complete convergence.
$(\Omega, \mathscr{F}, \mathsf{P})$	Complete probability space.
CADD	Conditional average detection delay.
$\mathsf{E}[X \mid \mathscr{B}]$	Conditional expectation of the random variable X given sigma-algebra $\mathscr{B}$.
$X_t \xrightarrow[t\to\infty]{\text{law}} Y$	Convergence in distribution (or in law or weak).
$X_t \xrightarrow[t\to\infty]{\mathsf{P}} Y$	Convergence in probability.
$F(x) = \mathsf{P}(X \le x)$	Cumulative distribution function (cdf) of a random variable X.
$\det A$	Determinant of the matrix A.
δ	Decision rule, procedure, function.
$\|\mathbf{X}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$	Euclidean norm.
E	Expectation.
ESS	Expected sample size (or average sample number).
$\mathsf{Expon}(\theta)$	Exponential distribution (or random variable) with the parameter θ.
$\{\mathscr{F}_t\}$	Filtration (a flow of sub-sigma-algebras $\mathscr{F}_t$).
$\dot{g}(x)$	First derivative of the function $x \mapsto g(x)$.
$\mathcal{F}$	Fisher information.
∇	Gradient (vector of first partial derivatives).
∇^2	Hessian (matrix of second partial derivatives).
$\mathbb{I}_n$	Identity matrix of size $n \times n$.
H_i	i^{th} hypothesis, $0 \le i \le M-1$, where M is the total number of hypotheses.
$\mathbb{1}_{\{A\}}$	Indicator of a set A.
A^{-1}	Inverse of the matrix A.
$\ker A$	Kernel of the matrix A.

I	Kullback–Leibler (K–L) information (or distance or divergence).
$\mathbb{R}^\ell$	ℓ-dimensional Euclidean space.
$\Lambda = \frac{d\mathsf{P}}{d\mathsf{Q}}(\omega)$	Likelihood ratio (Radon–Nikodým derivative of measure P with respect to measure Q).
$\varkappa$	Limiting average overshoot.
$L(\theta, d)$	Loss function.
$X_t \xrightarrow[t\to\infty]{L^p} Y$	L^p-convergence (or in the p^{th} mean).
$A = [a_{ij}]$	Matrix A of size $m \times n$ ($1 \le i \le m,\ 1 \le j \le n$).
$\mathbb{R}_+ = [0, \infty)$	Nonnegative real line.
X_t, $t \ge 0$	Observed process in continuous time.
X_n, $n \ge 1$	Observations in discrete time.
(a, b)	Open interval.
$\mathcal{P} = \{\mathsf{P}_\theta\}_{\theta\in\Theta}$	Parametric family of probability distributions.
$f_\theta(x)$, $p_\theta(x)$	Parametrized probability density, pdf.
P	Probability measure.
$f(x)$, $p(x)$	Probability density function (pdf).
θ	Parameter or vector of parameters.
ν	Point of change (or changepoint).
$\chi^2_m(p)$	p-quantile of the standard chi-squared distribution with m degrees of freedom.
β	Power of test.
α_{ij}	Probability of accepting H_i when the hypothesis H_j is true.
α_i	Probability of rejecting H_i when it is true.
$\mathsf{rank}\, A$	Rank of the matrix A.
$\mathbb{R} = (-\infty, \infty)$	Real line.
$X_t \xrightarrow[t\to\infty]{r-\text{quickly}} Y$	r-quick convergence.
$\ddot{g}(x)$	Second derivative of the function $x \mapsto g(x)$.
$\delta = (T, d)$	Sequential test (more generally rule).
$\mathbb{Z}_+ = \{0, 1, 2, \dots\}$	Set of nonnegative integers.
Ω	Set of elementary events ω.
$\{t : \dots\}$	Set of t such that $\dots$.
$\mathscr{F}$	Sigma algebra (field).
$\varphi(x)$	Standard normal density function.
$\Phi(x)$	Standard normal distribution function.
$\mathcal{N}(0,1)$	Standard normal random variable.
STADD	Stationary average detection delay.
$(\Omega, \mathscr{F}_t, \mathscr{F}, \mathsf{P})$	Stochastic basis.
T	Stopping time.
SADD	Supremum average detection delay.
d	Terminal decision.
$\mathbf{X}_0^t = \{X_u, 0 \le u \le t\}$	Trajectory of a random process observed on the interval $[0, t]$.
$A^\top$	Transpose of the matrix A.

$\mathrm{tr}\, A$	Trace of the matrix A.
$\mathbf{X}_1^n = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$	Vector of observed n random variables.
$\check{\mathbf{X}}_1^n = \begin{pmatrix} X_n \\ X_{n-1} \\ \vdots \\ X_1 \end{pmatrix}$	Vector of observed n random variables in reverse order.

Chapter 1

Motivation for the Sequential Approach and Selected Applications

In this chapter, we describe the theoretical and applied motivations for the sequential approach in general and for change detection in particular, and we describe the positioning of the book as well. We also introduce several typical application examples.

1.1 Motivation

Sequential analysis refers to statistical theory and methods for processing data in which the total number of observations is not fixed in advance but depends somehow on the observed data as they become available. A sequential method is characterized by two components:

1. A stopping rule that decides whether to stop the observation process with $(X_1, X_2, \ldots, X_n)$ or to get an additional observation X_{n+1} for $n \geq 1$;
2. A decision rule that specifies the action to be taken about the considered problem (estimation, detection, classification, *etc.*) after the observation has stopped.

Denoting by T the stopping variable and d the terminal decision, the pair $\delta \stackrel{\Delta}{=} (T,d)$ specifies the sequential decision rule (or procedure). Such a pair may not be unique for a given problem. The objective of sequential analysis is to determine an optimal decision rule δ that satisfies some criteria. Note that if T is fixed with probability 1 the procedure has an *a priori* fixed size of a sample. We will refer to such procedures as *Fixed Sample Size* procedures.

In sequential changepoint detection problems, however, the situation is slightly different. A change detection procedure is identified with a stopping time depending on the observations and the decision on no-change is equivalent to the decision on continuing observation. Furthermore, typically the observation process is not terminated even after deciding that the change is in effect but rather renewed all over again, leading to a multicyclic detection procedure. This is practically always the case in surveillance applications and often in other applications. See Section 6.3 for further details.

Even though most experiments are essentially sequential, many classical statistical methods are fixed sample size. In his history of sequential analysis, B.K. Ghosh distinguishes several practical motivations for sequential analysis [161].

In some applications sequential analysis is nothing but intrinsic: no fixed sample size procedure can be thought of. This is the case of industrial process control [81, 303, 482, 499, 501, 511]. This is also the case in the classical secretary problem [144] and while monitoring some critical health parameters of a patient in clinical trials [502]. Most surveillance problems are also sequential in nature. It should be noted that in the key area of medical and pharmaceutical research the requirement for sequential analysis may also result from ethical grounds.

In some other statistical inference applications, sequential analysis is the most economic solution, in terms of sample size or cost or duration of the experiment. This is the case of the so-called curtailed sampling procedure that ensures the same power while requiring a smaller sample size

than the best fixed sample size procedure [132, 189]. This is also the case of the repeated significance test that also maintains the flexibility of deciding sooner than the fixed sample size procedure at the price of some lower power [13, 514]. The sequential probability ratio test (SPRT) and the Kiefer–Weiss procedure also belong to the category of most economic solutions, since they minimize the expected sample size (resp. the maximum expected sample size). These sequential tests are investigated in detail in Chapters 3 and 5, respectively.

Finally, in some parametric sequential point estimation problems, sequential analysis may reinforce a fixed sample size procedure in a somewhat wider context than usual [311].

1.2 Two Theoretical Tracks

In this book we propose to focus on two tracks: Sequential Hypothesis Tests and Sequential (Quickest) Changepoint Detection.

First, classical settings of hypothesis testing and changepoint detection problems operate with the case of independent and identically distributed (iid) observations and two simple hypotheses. These assumptions may be quite restrictive for many contemporary applications. Therefore, generalizations to general non-iid models are under way. However, even in a relatively simple iid setting there are several challenges that have been addressed in the literature during the last decade, including the work by the authors. All these important results are scattered in the literature (conference proceedings as well as in statistical, applied probability, engineering, computer science, and other kinds of journals) and are not easily accessible and understandable for students and even for professionals in the field. Moreover, the practical needs of various applied areas lead the researchers to study more sophisticated statistical models by considering:

- Non-identically distributed and/or dependent observations,
- Multiple hypotheses,
- Composite hypotheses, including nuisance parameters in the statistical model.

Therefore, we believe that a book that would combine all these results in a synergistic way is timely.

Second, the proposed book contains both theoretical concepts and results and a number of application examples. As explained below and detailed in the table of contents, the book covers sequential hypothesis testing and sequential quickest changepoint detection from theoretical developments to applications in a wide range of engineering and environmental domains. It is the intention of the authors to explain how the theoretical aspects influence the problem statement and the design of algorithms when addressing problems in various application areas.

Third, we would like to mention two recent books related to sequential hypothesis tests and quickest change detection: by G. Peskir and A.N. Shiryaev, *Optimal Stopping and Free Boundary Problems* [360] and by H.V. Poor and O. Hadjiliadis, *Quickest Detection* [376]. While these books cover certain interesting aspects of sequential hypothesis testing and changepoint detection, they both focus mainly on continuous-time models, which are restricted for most applications. The present book covers mostly more practical discrete-time models as well as very general cases that include both continuous- and discrete-time models. In addition, we consider multiple decision making problems, including sequential multihypothesis tests and quickest change detection–isolation procedures, that are not presented in the above referenced books.

1.2.1 Track 1: Sequential Hypothesis Testing

The goal of testing statistical hypotheses is to relate an observed stochastic process to one of N ($N \geq 2$) possible classes based on some knowledge about the distributions of the observations under each class or hypothesis. In a sequential setting, the number of observations is allowed to be random, i.e., a function of the observations. The theoretical study of sequential hypothesis testing has been initiated by A. Wald [492]. A sequential procedure or test includes a stopping time and a terminal decision to achieve a tradeoff between the average observation time and the quality of the decision.

Most efforts have been devoted to testing two hypotheses, namely, to developing optimal strategies and obtaining lower bounds for the average number of observations necessary to decide between the two hypotheses with given error probabilities; see Wald [492, 494], Wolfowitz [496, 497], Hoeffding [192, 193], and many others. Also, these bounds have been compared with the sample size of the best non-sequential, fixed sample size test. It has been shown that the sequential procedure performs significantly better than the classical Neyman–Pearson test in the case of two simple hypotheses.

The problem of sequential testing of many hypotheses is substantially more difficult than that of testing two hypotheses. For multiple-decision testing problems, it is usually very difficult, if even possible, to obtain optimal solutions. The first results have been established by Sobel and Wald [435], Armitage [12], and Paulson [350]. The lower bounds for the average sample number has been established by Simons [432].

A substantial part of the development of sequential multihypothesis testing in the last several decades has been directed toward the study of suboptimal procedures, basically multihypothesis modifications of a sequential probability ratio test, for iid data models. See, e.g., Armitage [12], Chernoff [97], Dragalin [123], Dragalin and Novikov [127], Kiefer and Sacks [231], Lorden [269, 275], Pavlov [351, 352]. The generalization to the case of non-stationary processes with independent increments was made by Tartakovsky [449, 452, 457], Golubev and Khas'minskii [168], and Verdenskaya and Tartakovsky [484]. The condition of independence of the log-likelihood ratio increments was crucial in these works. Further generalizations to the case of non-iid stochastic models that may include both nonhomogeneous and correlated processes observed in continuous or in discrete time were made by Lai [248], Tartakovsky [455], and Dragalin *et al.* [128]. The results obtained in these latter works are indeed very general and cover almost any, and perhaps every, model of interest in the applications. Such popular models as Itô processes, state-space models, and hidden Markov models with discrete and continuous space are particular cases.

1.2.2 Track 2: Quickest Changepoint Detection

Changepoint problems deal with detecting changes in the state of a process. In the sequential setting, as long as the behavior of the observations is consistent with the initial or target state, one is content to let the process continue. If the state changes, then one is interested in detecting that a change is in effect, usually as soon as possible after its occurrence. Any detection policy may give rise to false alarms. The desire to detect a change quickly causes one to be trigger-happy, which will bring about many false alarms if there is no change. On the other hand, attempting to avoid false alarms too strenuously will lead to a long delay between the time of occurrence of a real change and its detection . The gist of the changepoint problem is to produce a detection policy that minimizes the average delay to detection subject to a bound on the average frequency of false alarms.

The theoretical study of quickest changepoint detection has been initiated in two different directions: Bayesian and minimax. In the Bayesian case, it is supposed that the changepoint is a random variable independent of the observations with known distribution. On the contrary, in the minimax case it is assumed that the changepoint is an unknown non-random number. The very first study of the Bayesian quickest changepoint detection approach has been done by Girschick and Rubin [165] in the framework of quality control. An optimal solution to this problem has been obtained by Shiryaev [413, 414, 415] who has also performed the comparison between the optimal procedure, the repeated sequential Wald test and the classical Neyman–Pearson test. Independently, another, minimax approach has been adopted by Lorden [271]. In contrast to the Bayesian approach, the minimax criterion is based on the worst-case mean detection delay, characterized by the essential supremum with respect to pre-change observations and by the supremum over all possible changepoints. An optimal solution to the problem and a lower bound in the class of procedures with a given mean time (average run length) to a false alarm has been studied by Lorden [271] in the asymptotic case for large average run length to false alarm. In this work, Lorden established, for the first time, asymptotic minimax optimality of Page's CUSUM procedure [346], a well-known statistical control

chart. Later Moustakides [305] showed that the CUSUM procedure is in fact exactly minimax with respect to Lorden's essential supremum detection speed measure.

In 1961, for detecting a change in the drift of a Brownian motion, Shiryaev [413, 414] introduced a change detection procedure, which is now usually referred to as the Shiryaev–Roberts procedure [394]. This procedure has a number of interesting optimality properties. In particular, it minimizes the *integral average detection delay* being Generalized Bayesian for an improper uniform prior distribution of the changepoint. It is also optimal in the sense of minimizing the stationary average detection delay when a change occurs in a distant future and is preceded by a long interval with a stationary flow of false alarms; see Feinberg and Shiryaev [139] and Pollak and Tartakovsky [370]. On the other hand, Pollak [365] introduced a natural worst-case detection delay measure — *maximal conditional average delay to detection*, which is less pessimistic than Lorden's essential supremum measure, and attempted to find an optimal procedure that would minimize this measure over procedures subject to constraint on the average run length to false alarm. Pollak's idea was to modify the Shiryaev–Roberts statistic by randomization of the initial condition in order to make it an equalizer. Pollak's version of the Shiryaev–Roberts procedure starts from a random point sampled from the quasi-stationary distribution of the Shiryaev–Roberts statistic. He proved that, for a large average run length to false alarm, this randomized procedure is asymptotically nearly minimax within an additive vanishing term. Since the Shiryaev–Roberts–Pollak procedure is an equalizer, it is tempting for one to conjecture that it may in fact be *strictly* optimal for any false alarm rate. However, a recent work of Moustakides *et al.* [310] and Polunchenko and Tartakovsky [373] indicates that the Shiryaev–Roberts–Pollak procedure is not exactly minimax and sheds light on this issue by considering a generalization of the Shiryaev–Roberts procedure that starts from a specially designed deterministic point.

As we mentioned above, in the early stages the theoretical development was focused on iid models. However, in practice the iid assumption may be too restrictive. The observations may be either non-identically distributed or correlated or both, i.e., non-iid. An extension of Lorden's results to the case of dependent stationary random processes before and after the change has been done by Bansal and Papantoni-Kazakos [26]. A general theory of changepoint detection is now available both in the Bayesian and minimax settings due to the work of Tartakovsky and Veeravalli [475, 476], Baron and Tartakovsky [28], Lai [251], and Fuh [154, 155]. In particular, for a low false alarm rate the asymptotic minimax optimality of the CUSUM and Shiryaev–Roberts procedures has been established in [154, 155, 251, 475] and the asymptotic optimality of the Bayesian Shiryaev procedure proven in [28, 476]. Moustakides [306] generalized for the Itô processes the CUSUM minimax optimality result with respect to Lorden's essential supremum measure acting on the total expected Kullback–Leibler information.

For iid data and for large thresholds, the suitably standardized distributions of the CUSUM and Shiryaev–Roberts stopping times are asymptotically exponential and fit well into the geometric distribution even for a very moderate false alarm rate [369]. In this case, the mean time to false alarm, the global false alarm rate metric, is obviously appropriate. However, for non-iid models the limiting distribution is not guaranteed to be exponential or even close to it. In general, we cannot even guarantee that large values of the mean time to false alarm will produce small values of the maximal local false alarm probability. Therefore, the mean time to false alarm, a standard and well accepted measure of false alarms, may not be appropriate in general. Instead of global measures of false alarms, it may be more appropriate to use local measures, for example the local false alarm probability, as suggested in [459]. This issue is extremely important for non-iid models as a discussion in [293, 460] and other discussion pieces published in *Sequential Analysis*, Vol. 27, No. 4, 2008 show.

Another challenging extension is a multidecision change *detection–isolation* problem when, along with detecting a change with a given false alarm rate, an identification/isolation of a true post-change hypothesis with a given misidentification rate is required [48, 49]. An optimal solution to the problem of abrupt change detection–isolation and a non-recursive algorithm that asymptotically attains the lower bound were obtained by Nikiforov in [322] by using a minimax approach

based on minimizing the Lorden-type worst-case mean detection–isolation delay for a given mean time before a false alarm and for a given probability of false isolation. The comparison between the optimal sequential and repeated fixed sample size approaches and different recursive sequential detection–isolation algorithms have been studied by Dragalin [125], Nikiforov [326, 328, 331], Oskiper and Poor [343], and Tartakovsky [453, 461]. A multiple hypothesis extension of the Shiryaev–Roberts procedure by adopting a dynamic programming approach has been proposed by Malladi and Speyer [287]. Next, Lai [252] generalized the results obtained for the worst-case mean detection–isolation criterion in [322] to the case of dependent observations. Lai also proposed two new optimality criteria: a non-Bayesian one, where the maximum probabilities of false alarm and false isolation within a given time window are constrained; and a Bayesian one, where a weighted sum of the false alarm and false isolation probabilities is used. Finally, Lai designed a window-limited generalized likelihood ratio-based algorithm with reduced computational complexity for on-line processing that asymptotically attains the lower bounds.

1.3 Several Applications

Hypothesis testing and changepoint problems arise across various branches of science and engineering and have an enormous spectrum of important applications, including environment surveillance and monitoring, biomedical signal and image processing, quality control engineering, link failure detection in communication networks, intrusion detection in computer networks and security systems, detection and tracking of covert hostile activities, chemical or biological warfare agent detection systems as a protection tool against terrorist attacks, detection of the onset of an epidemic, failure detection in manufacturing systems and large machines, target detection in surveillance systems, econometrics, financial markets, detection of signals with unknown arrival time in seismology, navigation, radar and sonar signal processing, speech segmentation, and the analysis of historical texts. In all of these applications, sensors take observations that undergo a change in their distribution in response to changes and anomalies in the environment or changes in the patterns of a certain behavior. The observations are obtained sequentially and, as long as their behavior is consistent with the normal state, one is content to let the process continue. If the state changes, then one is interested in detecting the change as soon as possible while minimizing false detections.

During the last years, a number of new application fields have emerged: structural health monitoring of bridges [24, 25, 43], wind turbines [178, 216], and aircraft [41, 102, 186, 188], detecting multiple sensor faults in an unmanned air vehicle (UAV) [403], monitoring railway vehicle dynamics [87], detecting road traffic incidents [521] or changes in highway traffic condition [170], monitoring low consumption components of road vehicles [36], diagnosing automotive antilock braking systems [285], chemical process control [196], physiological data analysis [398], surveillance of daily disease counts [439], nanoscale analysis of soft biomaterials through atomic force microscopy [402], biosurveillance [110, 342, 424], radio-astronomy [152, 438] and interferometry [341], spectrum sensing in cognitive radio systems [201, 263], landmine detection [379], leak detection in water channels [58], monitoring biological waste water treatment plants [19], environmental monitoring [57, 120, 361, 385, 409], hydrology [286], handling climate changes [284, 393, 526], navigation systems monitoring [295, 336, 408], detecting salient motion for dynamic scene modeling [233], human motion analysis [85], video scene analysis [262], sequential steganography [479, 480], biometric identification [7], onset detection in music signals [59], detecting changes in large payment card datasets [107], running consensus in sensor networks [82, 83], and distributed systems monitoring [382, 461, 475].

In particular a number of computer and network problems are now addressed with the aid of sequential hypothesis testing and change detection algorithms: anomaly detection in IP networks [477], secure IP telephony [386], detection of intrusion, viruses, and other denial of service (DoS) attacks [215, 357, 433, 472], including scanning worms infections [397, 406], bioterrorism detection and other aspects of global security, Internet access patterns characterization [208], teletraffic monitoring [2, 3, 211, 313], tracking the preferences of users in recommendation sys-

tems [520], network bandwith monitoring [183], active queue management [74], and even cost estimation for software evolution [383] and software quality and performance monitoring [171].

In this section, we describe several typical application examples of sequential hypothesis testing and change detection techniques. For each example, we give a short description of the particular problem and its context. For some of these models, the detailed information about the possibly complex underlying physical models is given in Part III. This selection of examples is not exclusive; it is intended to give only sufficient initial insights into the variety of problems that can be solved within this framework. In Part III, we come back to some application problems, showing results of the processing of real data with the aid of sequential hypothesis testing and change detection algorithms.

In Subsections 1.3.1 and 1.3.2 we start with quality control and target detection, and we continue with integrity monitoring of navigation systems in Subsection 1.3.3. Then in Subsection 1.3.4 we describe a couple of signal processing problems, namely segmentation of signals and seismic signal processing. Mechanical systems integrity monitoring is discussed in Subsection 1.3.5. Finally, we discuss application to finance and economics and to computer network surveillance and security in Subsections 1.3.6 and 1.3.7.

1.3.1 Quality Control

One of the earliest applications of change detection is the problem of quality control, or continuous production monitoring. On-line quality control deals with scenarios where the measurements are taken one at a time and the decisions are to be reached sequentially as the measurements are taken. Consider a production process that can be *in control* and *out of control*. The events associated with the transitions of this process from the in-control state to the out-of-control state are called *disorders*. For many reasons, it is necessary to detect a disorder as quickly as possible after its occurrence as well as to estimate its onset time. It may be a question of safety of the technological process, quality of the production, or classification of output production items. For all these problems, the best solution is the *quickest detection of the disorder with as few false alarms as possible*. This criterion is used because the delay until detection is a time interval during which the technological process is out of control, but there is no action of the monitoring system to this event. From both the safety and quality points of view, this situation is obviously highly undesirable. On the other hand, frequent false alarms are inconvenient because of the cost of stopping production, verifying whether this is a true or false disorder, and searching for the origin of the defect; nor is this situation desirable from a psychological point of view, because the operator will stop using the monitoring system very quickly if it produces too-frequent false alarms. Thus, an optimal solution is based on a *tradeoff* between the speed of detection or detection delay and the false alarm rate, using a comparison of the losses implied by the true and false detections.

We stress that we are interested in solving this problem using a *statistical approach*, that is, assuming that the measurements are a realization of a random process. Because of the random behavior, large fluctuations can occur in the measurements even when the process is in control, and these fluctuations result in false alarms. On the other hand, any (even the best) decision rule cannot detect the change instantaneously, again because of the random fluctuations in the measurements. When the technological process is in control, the measurements have a specific probability distribution. When the process is out of control, this distribution changes. If a parametric approach is used, we speak about changes in the parameters of this probability distribution. A chemical plant where the quality of the output material is characterized by the concentration of some chemical component is a typical example, where the concentration is distributed according to the Gaussian law. Under normal operating conditions, the mean value and standard deviation of this normal distribution are μ_0 and σ_0, respectively. Under abnormal conditions three types of changes can occur in these parameters:

- Deviation from the reference mean value μ_0 toward μ_1 with constant standard deviation, i.e., a systematic error;

- Increase in the standard deviation from σ_0 to σ_1 with constant mean, i.e., a random error;
- Both the mean and the standard deviation change, i.e., systematic and random errors.

The goal is to design a statistical decision rule (detection procedure, algorithm) that can detect these disorders effectively. Typically a decision procedure involves comparing a statistic sensitive to a change with a threshold that controls a false alarm rate.

If a decision statistic is chosen, the tuning of the statistical decision rule is reduced to selecting a threshold that guarantees the tradeoff between the false alarm rate and the mean delay to detection. Several types of decision rules are used in the industry as standards, they are called *control charts*, and each differs by the detection statistic. In the simplest case, the pre-change and post-change parameters are assumed to be known. In this case the decision statistics should be a function of the likelihood ratio for the pre- and post-change parameters.

The main references in the area of quality control and Statistical Process Control (SPC) are the books [80, 81, 114, 130, 153, 184, 288, 303, 340, 348, 434, 482, 499, 500, 501, 515] and the survey papers [65, 106, 443, 447, 509, 510, 511], with special notice for [381] and [67, 185].

1.3.2 Target Detection and Tracking

Surveillance systems, such as those for ballistic and cruise missile defense, deal with the detection and tracking of moving targets. The most challenging problem for such systems is the quick detection of maneuvering targets that appear and disappear at unknown points in time against a strong cluttered background. To illustrate the importance of this task, we remark that under certain conditions *a few seconds decrease* in the time it takes to detect a sea/surface skimming cruise missile can yield a significant increase in the *probability of raid annihilation*. Furthermore, usually detection systems are multichannel, since the target velocity is unknown. Thus, finding an optimal combination of a multihypothesis testing algorithm with changepoint detection methods is a challenge. This challenging applied problem can be effectively solved using the quickest detection–isolation methods developed in this book.

We also note that standard ad hoc methods for target track initiation and termination [27, 68, 69] can be substantially improved by using advanced quickest detection methods that are the subject of this book. Improving the operating characteristics is especially important for Space-Based Infrared and Space Tracking and Surveillance System sensors with chaotically vibrating lines-of-sight that have to provide early detection and tracking of low observable targets in the presence of highly-structured cluttered backgrounds.

1.3.3 Navigation System Integrity Monitoring

For many safety-critical aircraft navigation modes (landing, takeoff, *etc*.), a major problem of existing navigation systems consists in their lack of integrity. The integrity monitoring concept, defined by the International Civil Aviation Organization, requires a navigation system to detect the faults and remove them from the navigation solution before they sufficiently contaminate the output. Recent research shows that the quickest detection–isolation of the navigation message contamination is crucially important for the safety of the radio-navigation system, e.g., GPS, GLONASS, Galileo, *etc*. It is proposed *to encourage all the transportation modes to give attention to autonomous integrity monitoring of GPS signals* [93].

Monitoring the integrity of a navigation system can be reduced to a quickest change detection–isolation problem [21, 324, 325, 332]. The time when the fault occurs and the type of fault are not just unknown but sometimes can be intentionally chosen to maximize their negative impacts on the navigation system. Therefore, the optimality criterion should favor fast detection in the worst case with few false alarms and false isolations. Fast detection is necessary because abnormal measurements are taken in the navigation system between the changepoint (fault onset time) and its detection, which is clearly very undesirable. On the other hand, false alarms/isolations result in

lower accuracy of the estimates because incorrect information is used at certain time intervals. An optimal solution involves a tradeoff between these two contradictory requirements. The changepoint detection–isolation techniques developed in this book can be used for obtaining optimal solutions to this challenging problem. This is discussed in Section 11.1. Historical references related to inertial navigation system monitoring are [315, 506]. The integrity monitoring of navigation systems is investigated in [93, 227, 295, 324, 325, 332, 336, 446]. Some challenges are pointed out in [408].

1.3.4 Signal Processing Applications

1.3.4.1 Segmentation of Signals and Images

A first processing step of recognition-oriented signal processing consists in automatic segmentation of a signal. A segmentation algorithm splits the signal into homogeneous segments, with sizes adapted to the local characteristics of the analyzed signal. The homogeneity of a segment can be formulated in terms of the mean level or in terms of the spectral characteristics. The segmentation approach has proved useful for the automatic analysis of various biomedical signals, in particular electroencephalograms [11, 73, 78, 207, 213, 404] and electrocardiograms [172]. Several segmentation algorithms for recognition-oriented geophysical signal processing are discussed in [39]. A changepoint detection based segmentation algorithm has also been introduced as a powerful tool for the automatic analysis of continuous speech signals, both for recognition [10] and for coding [117].

The main desired properties of a segmentation algorithm are *low false alarm and mis-detection rates* and a *small detection delay*, as in the previous examples. However, we have to keep in mind that signal segmentation is usually only the first step of a recognition procedure. From this point of view, it is obvious that the properties of a given segmentation algorithm also depend upon the processing of the segments which is performed at the next stage. For example, it is often the case that, for segmentation algorithms, false alarms (sometimes called oversegmentation) are less critical than for onset detection algorithms. A false alarm for the detection of an imminent tsunami obviously has severe and costly practical consequences. On the other hand, in a recognition system, false alarms at the segmentation stage can often be easily recognized and filtered at the next stage, which means that the loss due to false alarms is small at the first segmentation stage. A segmentation algorithm exhibiting the above-mentioned properties is potentially a powerful tool for a recognition system.

It should be clear that a segmentation algorithm allows us to detect several types of events. Examples of events obtained through a spectral segmentation algorithm and concerning recognition-oriented speech processing are discussed in [10]. Other examples of events in seismology are mentioned in the previous subsection.

Changepoint detection methods are also efficient and useful in image segmentation and boundary tracking problems [96].

1.3.4.2 Seismic Data Processing

In many situations of seismic data processing, it is necessary to estimate *in situ* the geographical coordinates and other parameters of earthquakes.

The standard sensor equipment of a three-component seismic station results in the availability of records of seismograms with three components, namely the east-west, north-south, and vertical components. When an earthquake arises, the sensors begin to record several types of seismic waves (body and surface waves), among which the more important ones are the *P*-wave and the *S*-wave. The *P*-wave is polarized in the source-to-receiver direction, namely from the epicenter of the earthquake to the seismic station. Hence, it is possible to estimate the source-to-receiver azimuth α using the linear polarization of the *P*-wave in the direction of propagation of the seismic waves. The two main events to be detected are the *P*-wave and the *S*-wave; note that the *P*-wave can be very low-contrast with respect to seismic noise. The processing of these three-dimensional measurements can be split into three tasks:

1. On-line detection and identification of the seismic waves;

2. Off-line estimation of the onset times of these waves;
3. Off-line estimation of the azimuth using the correlation between the components of the *P*-wave segments.

The *P*-wave has to be detected *very quickly with a fixed false alarms rate*, so that the *S*-wave can also be detected on-line. The detection of the *P*-wave is a difficult problem, because the data contain many nuisance signals (interference) coming from the environment of the seismic station, and discriminating between these events and a true *P*-wave is not easy. The same is true for the *S*-wave, which is an even more difficult problem because of a low signal-to-noise ratio and numerous interferences between the *P*-wave and the *S*-wave.

After *P*-wave and *S*-wave detection, the *off-line accurate estimation of onset times* is required for both types of waves. A possible solution is to use fixed-size samples of the three-dimensional signals centered at a rough estimate of the onset time provided by the detection algorithm. Some references for seismic data processing are [235, 301, 334, 363, 377, 478].

1.3.5 Mechanical Systems Integrity Monitoring

Detecting and localizing damages for monitoring the integrity of structural and mechanical systems is a topic of growing interest, due to the aging of many engineering constructions and machines and to increased safety norms. Many structures to be monitored, e.g., civil engineering structures subject to wind and earthquakes, aircraft subject to turbulence, are subject to both fast and unmeasured variations in their environment and small slow variations in their modal or vibrating properties. While any change in the excitation is meaningless, damages or fatigues on the structure are of interest. But the available measurements do not separate the effects of the external forces from the effect of the structure. Moreover, the changes of interest, that may be as small as 1% in the eigenfrequencies, are visible neither on the signals nor on their spectra. A global health monitoring method must rather rely on a model which will help in discriminating between the two mixed causes of the changes that are contained in the data. This vibration monitoring problem can be stated as the problem of detecting changes in the autoregressive (AR) part of a multivariable autoregressive moving average (ARMA) model having nonstationary MA coefficients. Change detection turns out to be very useful for this monitoring purpose, for example for monitoring the integrity of the civil infrastructure [24, 25, 45].

The improved safety and performance of aerospace structures and reduced aircraft development and operating costs are major concerns. One of the critical objectives is to ensure that the newly designed aircraft is stable throughout its operating range. A critical aircraft instability phenomenon, known as flutter, results from an unfavorable interaction of aerodynamic, elastic, and inertial forces, and may cause major failures. A careful exploration of the dynamical behavior of the structure subject to vibration and aeroservoelastic forces is thus required. A major challenge is the in-flight use of flight test data. The flight flutter monitoring problem can be addressed on-line as the problem of detecting that some instability indicators decrease below some critical value. CUSUM-type change detection algorithms are useful solutions to these problems [41, 46, 296, 531].

These application examples illustrate change detection with estimating functions different from the likelihood [36, 38].

The vibration-based structural health monitoring problem is explored in Section 11.2.

1.3.6 Finance and Economics

Stochastic modeling in finance is a new application area for optimal stopping and quickest change-point detection. For example, in the *Russian option* [410] the fluctuations in the price of an asset are modeled by geometric Brownian motion (the Black–Sholtz model), and the problem consists in finding a stopping time that maximizes a certain gain. In this optimization problem, the option owner is trying to find an exercise strategy that maximizes the expected value of his future reward

with a certain interest rate for discounting. This problem can be effectively solved using the optimal stopping theory which is a part of the book. A similar approach can be applied for finding an optimal solution to the *American put option* with infinite horizon [359].

An application of the optimal stopping theory in financial engineering imposes an analysis for the gain process depending on the future and referring to an optimal prediction problem, which falls outside the scope of the classical optimal stopping framework. A typical setting is related to minimizing over a stopping time a functional of a Brownian motion.

These examples show that the optimal stopping theory can be effectively applied to many probabilistic settings of theoretical and practical interest. In addition, we mention the articles [52, 358] and references therein.

We also argue that quickest changepoint detection schemes can be effectively applied to the analysis of financial data. In particular, quickest changepoint detection problems are naturally associated with rapid detection of the appearance of an arbitrage in a market [421].

1.3.7 Computer Network Surveillance and Security

A considerable interest exhibited over the past decade in the field of defense against cyber-terrorism in general, and network security in particular, has been induced by a series of external and internal attacks on public, private corporate, and governmental computer network resources. Malicious intrusion attempts occur every day and have become a common phenomenon in contemporary computer networks. Examples of malicious activities are spam campaigns, phishing, personal data theft, worms, distributed denial-of-service (DDoS) attacks, address resolution protocol man-in-the-middle (ARP MiM) attacks, fast flux, *etc.* These pose an enormous risk to the users for a multitude of reasons such as significant financial damage, or severe threat to the integrity of personal information. It is therefore essential to devise automated techniques to detect such events as quickly as possible so that an appropriate response can be provided and the negative consequences for the user can be eliminated.

The detection of traffic anomalies is done by employing an intrusion detection system (IDS). Such systems in one way or another capitalize on the fact that malicious traffic is noticeably different from legitimate traffic. Depending on the principle of operation there are two categories of IDSs: either signature or anomaly based [113, 224]. A signature-based IDS inspects the passing traffic with the intent to find matches against already known malicious patterns. By contrast, an anomaly-based IDS is first trained to recognized the normal network behavior and then watches for any deviation from the normal profile.

Currently both types of IDSs are plagued by a high rate of false positives and the susceptibility to carefully crafted attacks that blend themselves into normal traffic. These two systems are complementary, and neither alone is sufficient to detect and isolate the myriad of network malicious or legitimate anomalies generated by attacks or other non-malicious events.

Intrusions usually lead to an abrupt change in the statistical characteristics of the observed traffic. For example, DDoS attacks lead to changes in the average number of packets sent through the victim's link per unit time. It is therefore appealing to formulate the problem of detecting computer intrusions as a *quickest changepoint detection problem*: to detect changes in statistical models as rapidly as possible, i.e., with minimal average delays, while maintaining the false alarm rate at a given low level. The feasibility of this approach has been already demonstrated in [472, 473, 474].

To make the detection delay small one has to increase the false alarm rate (FAR), and *vice versa.* As a result, the FAR cannot be made arbitrarily low without sacrificing other important performance metrics such as the detection delay and the probability of detection in a given time interval. Therefore, while attack detection algorithms can run with very low delay, this comes at the expense of high FAR, and thus changepoint detection techniques may not be efficient enough for intrusion detection. The ability of changepoint detection techniques to run at high speeds and with low delay, combined with the generally low frequency of intrusion attempts, presents an interesting opportunity: What if one could combine such techniques with others that offer very low false alarm rates

but are too heavy to use at line speeds? Do such synergistic IDSs exist, and how can they be integrated? Such an approach is explored in Section 11.3. Specifically, a novel hybrid approach to network intrusion detection that combines *changepoint detection based anomaly* IDS with a *flow-based signature* IDS is proposed. The proposed hybrid IDS with profiling capability complements existing anomaly- and signature-based systems. In addition to achieving high performance in terms of the tradeoff between delay to detection, correct detection, and false alarms, the system also allows for isolating the anomalies. Therefore, the proposed approach overcomes common drawbacks and technological barriers of existing anomaly and signature IDSs by combining statistical changepoint detection and signal processing methods.