Event Correlation Using Conditional Exponential Models with Tolerant Pattern Matching Applied to Incident Detection

A Thesis presented to the
Fachbereich Mathematik/Informatik
Universität Bremen
In Partial Fulfillment of the Requirements for the Degree
Dr.-Ing.

Dissertation

by
Carsten Elfers
born in Gronau

Submission: December, 2013
Colloquium: September, 2014

Advisor: Prof. Dr. Stefan Edelkamp
Co-Advisor: Prof. Dr. Norbert Pohlmann

This revised version includes minor corrections of the layout, a shortened appendix, minor clarifications and updates. This version comes without digital attachment.

Berichte aus der Informatik

Carsten Elfers

Event Correlation Using Conditional Exponential Models with Tolerant Pattern Matching Applied to Incident Detection

D 46 (Diss. Universität Bremen)

Shaker Verlag
Aachen 2014

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Zugl.: Bremen, Univ., Diss., 2014

Printed in Germany.

ISBN 978-3-8440-3168-3
ISSN 0945-0807

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen
Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9
Internet: www.shaker.de • e-mail: info@shaker.de

Thanks to...

- **Bettina Elfers**
  my wife, for making square things round.

- **Hartmut Messerschmidt**
  a colleague and friend, for his help in the analysis of the Conditional Exponential Model and his great — and funny — lectures about machine learning.

- **Heinz and Maria Elfers**
  my parents, for always being there for me.

- **Jennifer Santa Lopes**
  my sister-in-law, for spell-checking this thesis.

- **Karsten Sohr**
  a colleague and friend, for his advises and his exhilarating and motivating work in the fides project.

- **Norbert Pohlmann**
  my co-advisor, for his support.

- **Otthein Herzog**
  for supporting me and my work.

- **Stefan Edelkamp**
  my thesis advisor, for encouraging discussions, supporting me and my thesis and his help with the complexity analysis of the Pareto algorithm.

Contents

Notation

In this work $\lg n$ refers to the dual logarithm $\log_2 n$, while $\ln n$ refers to the natural logarithm $\log_e n$. Vectors or sets are denoted by bold characters, e.g. $\boldsymbol{x}$. Elements are indexed by a subscript enumerator, e.g., x_j and are written in bold characters if they are vectors as well, for example, $\boldsymbol{x}_j$. A non-bold character without index, e.g., x, denotes an arbitrary element of $\boldsymbol{x}$. The power set is denoted by $\mathcal{P}$. The following table gives an overview of the most frequently used symbols in this work (sorted by Latin letters before Greek letters):

Symbol	Description
F	Fusion function
j, k, l, m, n	These symbols are reserved for indexing elements.
$\boldsymbol{e}$	Set of entities used in the constraints
$\boldsymbol{f}$	Set of feature functions
$\boldsymbol{h}$	Set of hypotheses
$\boldsymbol{i}$	Set of incidents
$\boldsymbol{p}$	Set of patterns
$\boldsymbol{q}$	Abstraction levels in the search space of the Pareto algorithm
$\boldsymbol{s}$	Set of samples
$\boldsymbol{t}$	Set of threat levels
$\boldsymbol{v}$	Set of pattern matching values
$\boldsymbol{x}$	Sequence of observations
$\boldsymbol{y}$	Sequence of labels
$\boldsymbol{R}$	Set of relations in the ontology
$\boldsymbol{S}$	Set of solutions of the Pareto algorithm
$\boldsymbol{Q}$	Search space of the Pareto algorithm
β	Penalty factor of the Fusion Function
θ	Similarity function
$\boldsymbol{\gamma}$	Set of constraints
$\boldsymbol{\lambda}$	Set of model parameters (weights)

Table 1: Table of symbols.