



Proceedings of

CERC2012

Collaborative European Research Conference

26 – 27 April 2012
Darmstadt, Germany

Editors

Patrick Bours
Bernhard Humm
Robert Loew
Ingo Stengel
Paul Walsh

Cover Pictures © Ulrich Mathias
Mathildenhöhe, Wissenschaftsstadt Darmstadt

CERC2012

Darmstadt, Germany
26 – 27 April 2012

Proceedings of the
**Collaborative European
Research Conference 2012**

Editors :
Patrick Bours
Bernhard Humm
Robert Loew
Ingo Stengel
Paul Walsh

ISSN : 2220 - 4164

© 2012 Hochschule Darmstadt – University of Applied Sciences
All rights reserved

Cover Pictures © Ulrich Mathias
Mathildenhöhe, Wissenschaftsstadt Darmstadt

No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means – electronic, mechanical, photocopy, recording or otherwise, without the prior written permission of the publisher or distributor.

Table of Content

Chapter 1

Computer Science

Telephony Fraud Detection in Next Generation Networks.....	13
S. Augustin, C. Gaißer, J. Knauer, M. Massoth, K. Piejko, D. Rihm, T. Wiens	
A Criteria-Driven Method for Architecting Domain-Specific IE Applications.....	21
T. Neef, S. Morana, B. G. Humm	
Attacks to ZigBee and Wireless Sensor Networks – Honeypots for Detection and Response.....	29
J. Markert, M. Massoth, K.-P. Fischer-Hellmann, S. M. Furnell	
Development of Three-Dimensional User Interfaces based on Low-Cost Inertial Navigation Systems	37
M. Muentler, M. Haid, T. Chobtrong, E. Guenes, M. Kamil	

Chapter 2

Media

Demoscene Computer Artists and Community.....	43
C. Hastik, A. Steinmetz	
Cognitive prototypes and narrative thinking.....	49
P. Green	

Chapter 3

Business

Managing the Future Energy Policy for Ireland : Examining the Role of Nuclear Power.....	59
D. Lynch, A. Wright	
Electric Vehicles in Ireland : The Future ?.....	69
N. O'Mahony, A. Wright	
Forecasting an EU farm gate milk price using the system dynamics methodology.....	75
D. Bergmann	
Political Marketing Segmentation within English Local Government Elections.....	83
R. L. Tidy	

Chapter 4

Electronic Engineering

Overall Transparency in Distribution Planning using Software-based Solutions and Low-cost Inertial Navigation Systems.....95
M. Haid, M. Kamil, T. Chobtrong, E. Günes, M. Münter, H. Tutsch

Bolt-Identification using an IMU with Bayesian Decision Theory.....101
T. Chobtrong, M. Haid, E. Günes, M. Kamil, M. Münter

Chapter 5

Electrical Engineering

New approaches for energy optimisation in Smart Homes and Smart Grids by automated service generation and user integration.....111
M. Steinheimer, U. Trick, P. Ruhrig

Simulation and Experimental Validation of a Pseudonoise Method for Multi-Fault Location and Identification in the presence of Noise on Transmission Line Systems.....121
R. A. Guinee, C. Healy

Chapter 6

Civil Engineering

Response of Precast Prestressed Concrete Circular Tanks Retaining Heated Liquids.....131
M. J. Minehane, B. D. O'Rourke

The Development of a Low Cost Instrument for the Measurement of Tidal Stream and Run of River Flows.....141
T. Daly, C. Gibbons, D. O'Reilly

Geo-spatial Data Fusion for Anomaly Detection.....151
J. Corkish, P. Walsh

Preface

Interdisciplinary Collaboration is a source of innovation and research. The Collaborative European Research Conference (CERC 2012) is an event to foster collaboration among friends and colleagues across disciplines and nations within Europe. Emerged from a long-standing cooperation between the Cork Institute of Technology, Ireland and Darmstadt University of Applied Sciences, Germany, CERC has this year been extended to include more well-established partners: Plymouth University, UK and Gjøvik University College, Norway.

CERC is truly interdisciplinary, bringing together young researches from science, engineering, business, humanities, and the arts. CERC is innovative, also in the way of creating a real workshop atmosphere. Researchers not only present their findings as published in their research papers. They are also challenged to collaboratively work out joint aspects of their research within conference sessions. Highlights will be presented in a plenary session and the best session presentation will be awarded a price.

To organize such an event involves the hard work of a number of people. Thanks go to the international program committee and my fellow program chairs, particularly to Dr Ingo Stengel for organizing the review process. Dr Robert Loew put a lot of work into preparing the proceedings. He, Janina Fengel, and Prof Udo Bleimann were invaluable for local organization. Thanks also to Prof Bernd Steffensen for supporting CERC, also financially.

Prof. Dr. Bernhard Humm

General Conference & Programme Co-Chair, CERC2012

Darmstadt, April 2012

CERC2012 Committees

General Conference & Programme Co-Chairs

Prof. Dr. Bernhard Humm
University of Applied Sciences Darmstadt, Germany

Dr. Paul Walsh
Cork Institute of Technology, Ireland

Dr. Ingo Stengel
Plymouth University, United Kingdom

Dr. Patrick Bours
University College Gjøvig, Norway

Organizing Committee

Udo Bleimann

Janina Fengel

Bernhard Humm

Robert Loew

Ingo Stengel

International Programme Committee

Shirley Atkinson
Plymouth University, United Kingdom

Udo Bleimann
University of Applied Sciences Darmstadt, Germany

Patrick Bours
University College Gjøvig, Norway

Shun-Ping Chen
University of Applied Sciences Darmstadt, Germany

Nathan Clarke
Plymouth University, United Kingdom

Barry O'Connor
Institute of Technology Cork, Ireland

Tom O'Connor
Institute of Technology Cork, Ireland

John Creagh
Institute of Technology Cork, Ireland

Paul Dowland
Plymouth University, United Kingdom

Klaus-Peter Fischer-Hellmann
University of Applied Sciences Darmstadt, Germany

Orla Flynn
Institute of Technology Cork, Ireland

Steven Furnell
Plymouth University, United Kingdom

Johannes Gerdes
University of Applied Sciences Darmstadt, Germany

Paul Green
Institute of Technology Cork, Ireland

Angelika Groterath
University of Applied Sciences Darmstadt, Germany

Richard Guinee
Institute of Technology Cork, Ireland

Klaus Habermehl
University of Applied Sciences Darmstadt, Germany

Bernhard Humm
University of Applied Sciences Darmstadt, Germany

CERC2012 Committees

Matthias Knoll

University of Applied Sciences Darmstadt, Germany

Andrea Krajewski

University of Applied Sciences Darmstadt, Germany

Manfred Loch

University of Applied Sciences Darmstadt, Germany

Robert Loew

University of Applied Sciences Darmstadt, Germany

Michael Loftus

Institute of Technology Cork, Ireland

Jonathan Moizer

Plymouth University, United Kingdom

Mark Phillips

Plymouth University, United Kingdom

Werner Sanns

University of Applied Sciences Darmstadt, Germany

Arnd Steinmetz

University of Applied Sciences Darmstadt, Germany

Ingo Stengel

Plymouth University, United Kingdom

Andreas Thuemmel

University of Applied Sciences Darmstadt, Germany

Alexander Vogel

University of Applied Sciences Darmstadt, Germany

Paul Walsh

Institute of Technology Cork, Ireland

Angela Wright

Institute of Technology Cork, Ireland

Chapter 1

Computer Science

Telephony Fraud Detection in Next Generation Networks

S. Augustin, C. Gaißer, J. Knauer, M. Massoth, K. Piejko, D. Rihm, T. Wiens

Department of Computer Science
Hochschule Darmstadt – University of Applied Sciences, Darmstadt, Germany
e-mail: michael.massoth@h-da.de

Abstract

Telephony fraud is a growing problem for telecommunication service providers that operate Next Generation Networks (NGN). This paper describes a framework for a rule-based fraud detection system. The classification of fraudulent calls is based on Call Detail Records (CDR) that are used by telecommunication service providers for billing purposes. By analysing this data, fraud can be detected efficiently. We propose a method for accomplishing this. The work has been conducted in collaboration with a telephony service provider that made real-life CDR data available for analysis. The main achievement of this paper is the description of a rule-based system that detects telephony fraud using CDR data.

Keywords

Communication system security, Communication system signalling, Communication system traffic, Computer network management, Next generation networking

1. Introduction

Telephony fraud is a serious problem for carriers that operate Next Generation Networks (NGN). Attackers regularly try to compromise accounts of users or providers to circumvent charging systems or to cause financial harm to customers. Telephony fraud comprises unauthorized deletion or alteration of billing records, unauthorized bypassing of lawful billing systems, unauthorized billing and the taking of service provider property (Zar, 2005).

1.1. Current situation

The Communications Fraud Control Association (CFCA) estimated in 2009 that fraud leads to a worldwide annual loss of 74 to 80 billion USD (Communications Fraud Control Association, 2009). It is expected that this value will increase in the future. The top three fraud types, as named in their report, are (see Figure 1):

- Subscription or identity theft (22.0 billion USD)
- Compromised Private Branch Exchange (PBX) systems (15.0 billion USD)
- Premium rate service fraud (4.5 billion USD)

Even single fraud attacks may cause significant losses. In one case, an attacker conducted 11,000 calls to Australia, causing an estimated damage of more than 120,000 USD. These calls were made over a period of only 46 hours (Tindal, 2009). These losses could be drastically reduced if effective real-time fraud detection mechanisms were applied.

This kind of fraud also causes significant economic damage because some small- and medium-sized enterprises (SME) may not be able to deal with the amount of financial damage caused by these attackers, possibly leading to bankruptcy.

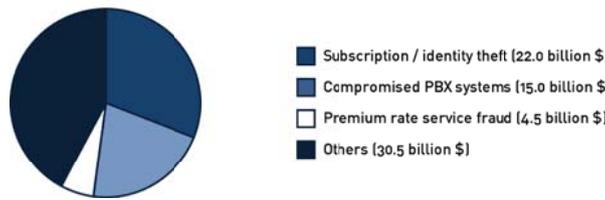


Figure 1: Top three fraud types

1.2. Challenges in fraud detection

In order to develop well performing fraud detection mechanisms, access to real world data is necessary. However, telecommunication providers are not allowed to expose this data due to privacy reasons. This is caused by national legal limitations, for example the German “Bundesdatenschutzgesetz” (Federal Data Protection Act) (Bundesministerium der Justiz, 2009). Additionally, fraud detection is not just a binary problem. The precise classification of calls as fraudulent or not with a minimum of false positives is difficult. There are cases that cannot be decided with certainty. Therefore, fraud detection has to be treated as an n-class problem (Padmaja, 2007).

1.3. Structure of the paper

This paper is structured as follows: Section 2 gives an overview on the recent activities in the field of fraud detection. Section 3 describes the basic concept of fraud detection and our design decisions for the framework. After the fundamentals have been explained, a more detailed description of our approach is given in Section 4. The paper ends with a conclusion and an outlook on future work in Section 6. Acknowledgements follow in the last section.

2. Related work

In this paper, a rule-based system for fraud detection is described. The field of fraud detection can be divided into multiple categories. Two important ones are rule-based approaches and neural networks. There are also additional approaches, for example Bayesian Networks, Support Vector Machines and Hidden Markov Models. These are described in Section 2.3 (see Figure 2).

2.1. Rule-based methods

Rule-based methods are very effective, but hard to manage. Extensive work is required to specify rules for every imaginable fraud case. Another downside is that rule-based fraud detection systems need to be updated frequently to cover new kinds of fraud (You et al., 2004).

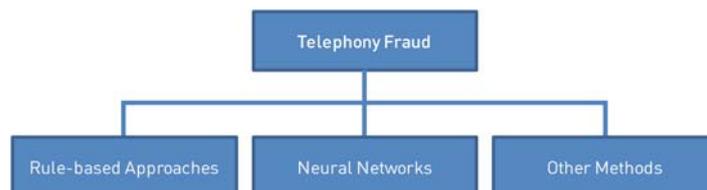


Figure 2: Methods to counter telephony fraud

Rosset et al. (Rosset et al., 1999) proposed an extension of the C4.5 algorithm, which is a popular algorithm to generate decision trees. The extension divides a rule-discovery process into two steps. The first step generates a large number of candidate rules. The second step puts together a rule-set from these candidates. Olszewski (Olszewski, 2011) constructed a detection method based on user profiling by employing the Latent Dirichlet Allocation (LDA). Using the Kullback-Leibler divergence, the participants are classified as “good” or “evil”. Ruiz-Agundez et al. (Ruiz-Agundez et al., 2010) propose an architecture for rule-based mechanisms that can be applied in NGN infrastructures.

2.2. Neural networks

One of the alternatives to rule-based approaches for classification are neural networks. These are more suitable to cover new and unknown attacks. Taniguchi et al. (Taniguchi et al., 1998) summarize three methods for fraud detection, one being a neural network. They claim that these three types are able to detect 85% of all fraud cases that occurred in their test set.

1. The first method consists of the application of a feed-forward neural network. It is used to learn a discriminative function to classify service subscribers using summary statistics.
2. The second method applies a Gaussian mixture model to determine the probability of the user's future behavior. This is based on user behavior in the past. The probabilities are used to validate the current behavior in order to detect deviations.
3. The third method uses a Bayesian network. Here, statistical properties of users and of multiple fraud cases are used.

The application of neural networks for fraud detection in mobile communication has been introduced by Qayyum et al. (Qayyum, 2010). A disadvantage of their approach is that further adjustments are needed for the system in order to work efficiently.

2.3. Other methods

The pattern recognition skills of the human eye are very powerful. Therefore, Cox et al. (Cox et al., 1997) proposed to apply humans in the process of fraud detection. They introduced multiple techniques to visualize network traffic in a human readable way. Hollmén and Tresp (Hollmén and Tresp, 1999) proposed a system that is based on a hierarchical regime-switching model. This system receives inference rules from a junction tree algorithm and is trained by using the Expectation Maximization (EM) algorithm.

3. Concept and overall system design

Every internet telecommunication service provider uses charging systems that log each call that was made using the network of the service provider. These log files contain detailed information about calls, and are commonly referred to as Call Detail Records, or sometimes as Call Data Records (CDR). In the CDR, the subscriber numbers of caller and callee, the date and time when the call was made and the call duration are recorded. Therefore, these log files contain valuable information that can be used to detect telephony fraud. Since CDR data is not allowed to be exposed to the public because of German legal regulations, the data provided by the cooperating telecommunication service provider had to be anonymised.

Our system uses CDR files and analyses them for anomalies (see Figure 3). This is accomplished by different filters. Each filter scans the CDRs using specific rules. If an anomaly is detected, and one of the filters supplies a positive result, there is a strong suspicion that a fraud case has occurred. This fraud case has to be validated by a human and further actions, for example the temporary deactivation of an account, have to be taken. Our framework does not automatically perform these actions, as telephony fraud comprises false positives.

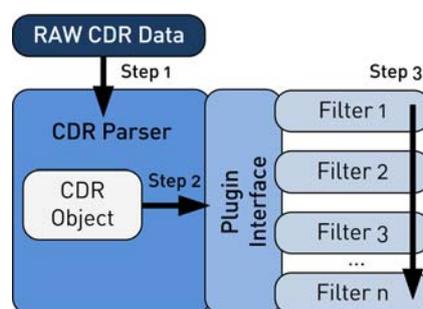


Figure 3: System overview

The framework has been implemented in Python 2.7. The decision to use Python resulted from several considerations. First of all, Python can be learned quickly and, due to its code structure, is easy to read. This ensures a quick start of implementation and results in low costs for later maintenance and the addition of extensions. Furthermore, Python is an open source product that is highly portable and runs on almost every operating system (P. S. Foundation, 2011).

4. System components

In this section, the system components are described in detail.

4.1. Structure of a Call Detail Record

Each CDR consists of several elements that correspond to different functionalities. These elements indicate the start and the end of a call, among other parameters. Each element contains the date and time when the element was written. The first element, indicating the beginning of a record, contains the unique session ID that identifies a CDR. The elements that are necessary for further analysis are now described in more detail.

The Incoming element of a CDR (called A-element in the CDR specification) contains the properties of an incoming call (TELES, 2006). For our purpose, only the carrier ID (the n-attribute of the A-element) is important.

The Connected element (C-element) only exists if a conversation was established. The C-element consists of several sub-elements. For example, its x-element contains the Session Initiation Protocol-(SIP) data of the connection. The SIP data contains several fields. The first field corresponds to the number of the callee. The 13th and 25th field both contain the customer ID or the subscriber number. Furthermore, the C-element includes the duration of a call in milliseconds.

If a call lasts longer than 15 minutes, the CDR is split into multiple parts. These parts can be identified by the first number in the S-element. This element is the first element in a CDR, indicating the beginning of the CDR. If the call duration is below 15 minutes, the identifier is set to "0". If it indicates the start of a record series, it is set to "1". The final part is marked as "3". All parts in between are set to "2".

If a call is finished, the Disconnecting element (D-element) is written. In this element, the reason for the call's termination is stored. The From-field in this element is also important, as it indicates which party hung up. In a nutshell, the C- and the D-element provide the necessary information to bill a call.

4.2. Framework

To analyse the CDRs, we developed a framework that is capable of parsing the log files generated by the billing system. The framework consists of multiple parts:

- Classes for CDRs and CDR-elements parsing the input data.
- The main part of the software controlling the application flow.
- Several filters implementing the rules for fraud detection.

Now, the individual parts of the framework are explained in more detail.

1. CDR Classes: The framework contains classes for each CDR element (see previous section). This modular structure provides easy filter access to the different CDR elements.
2. Main part: This part of the software controls the application flow. It starts the application, evaluates the console commands for the input files that are to be parsed and registers the different filters. The filters are organized as a list, which is iterated for each input CDR. To expand the software, more filters can easily be integrated into the analysis process, simply by adding them to the list of registered filters.

The CDR parser starts to read the data from the given input files. Each CDR is parsed from the log files into a CDR object. Each filter expects a CDR object as input and analyses it. After the input files have been parsed completely, the results from the filters are collected by the main part. If one filter or multiple filters have detected a potential fraud case, the output is saved to a text file. The output contains each filter's result as a binary value. Then, an operator is alarmed.

The release candidate comes as a console application. A graphical user interface has not been included, since the software is used by the technical staff of the cooperating telecommunication service provider and the systems that process the CDRs are UNIX-based.

4.3. Filters

The framework includes a filter base class that is inherited by all implemented filters (see Figure 4). This base class contains methods for all filters, e.g., for the formatting of date and time, and a method that returns the results. For each rule, which was defined to detect fraud, a filter is implemented. Each filter analyses a given CDR, evaluates it for fraud-suspicious data and returns the collected results to the main class.

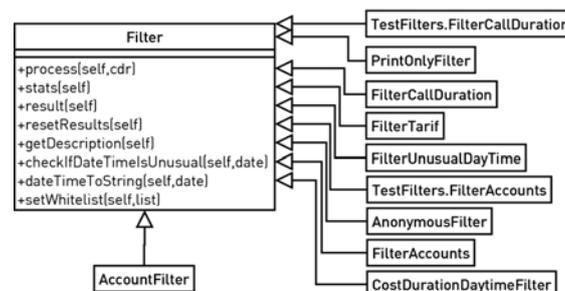


Figure 4: Filter base class and inherited classes

In general, all filters only regard calls originating from the internet telecommunication service provider's network, as only these calls are charged. These are identified if the callee's subscriber number corresponds to a customer ID and the carrier ID in the Incoming element of the CDR does not correspond to the service provider's ID.

Up to date, four filters have been developed. The first filter regards only single calls of a customer. The second one regards all calls of a specific customer per hour. The third filter scans for signalling errors and suppressed caller IDs, while filter number four considers historical user data.

The first filter analyses a single call for the following criteria:

- The duration of the call, depending on the destination pay scale area.
- The date and time when the call was made.

To classify the pay scale area, the destination area code of the callee's subscriber number is analysed. We defined four categories of pay scale areas:

1. No charges: The first category classifies calls that only cause low charges or none at all. Therefore, these calls are omitted. As the software was developed in cooperation with a German company, the relevant area codes include the German fixed network, Voice over Internet Protocol (VoIP) and national subscriber numbers.
2. Moderately expensive: This category comprises calls destined for the German mobile network. These calls are not very expensive, regarding the charges per minute. In this case, calls lasting for more than a specific threshold are considered unusual.
3. Expensive: To simplify the classification, this category includes all calls that do not belong to one of the other categories. These are calls that are destined for international and special rate numbers. A threshold for the call duration is set accordingly.
4. Very expensive: Satellite calls belong to the most expensive category. These calls may be charged at up to 20 per minute. Therefore, the threshold in this category is considerably lower than the thresholds in the previous categories. The second criteria for this filter are the date and the time when the call took place. If, for example, a company only has business customers, it can be assumed that calls outside the business hours or on weekends are more suspicious than others.

The second filter regards all calls that are made by a specific customer in a given time frame. The criteria are as follows: If the amount of calls per hour is greater than a specific value or if the overall call duration per hour exceeds a specific threshold, it is assumed that this is a fraudulent usage of the telephony service.

The first and the second filter also include a whitelist for specific customers. Whitelist candidates are customers who would regularly be above the thresholds with their normal call behaviour, and therefore would be considered as fraudulent. Those customers are maintained in the whitelist and are ignored by the filters.

The third filter scans the input data for signalling errors and suppressed caller IDs, since these may also denote fraud cases. These parameters are only considered for analysis if they are found on incoming calls. Additionally, data in the CDRs indicating the connection quality is assessed by this filter. One of the typical fraud scenarios consists of routing calls via multiple international service providers. In these cases, connection quality may drop significantly. Therefore, low connection quality may be another indicator for fraud cases.

The fourth filter collects historical user data, for example the total duration of calls made by a single user or by all users. Here, up to seven categories may be included. Additionally, this filter is able to output descriptive statistics and diagrams as a PDF file.

Another interesting information contained in a CDR is the reason for call termination. This is stored in the D-element. Among the possible reasons, SIP-signalling errors and identity errors are the most interesting ones from the perspective of fraud detection. These reasons can also be used for statistical purposes or to detect internal network errors.

The filter rules and their associated thresholds have been determined by a thorough evaluation of actual fraud cases. This has been actively supported by the collaborating service provider. Unfortunately, it is not possible to describe the rules and thresholds in more detail. A publication of these parameters would give attackers a significant advantage in bypassing the system, which is productively used.

5. Conclusion and future work

In general, the presented rule-based approach for detecting telephony fraud is promising. The described solution performs well on the real-life CDRs delivered by the service provider, regularly classifying about 4% as false positive fraud cases. Additionally, it is almost an order of magnitude faster than the solution previously used, which was script-based. For example, the presented system is able to process typical CDR files in significantly less than one minute, while the old system took more than ten minutes to accomplish this, under identical circumstances. Furthermore, the system did not only detect known fraud attacks, but also discovered yet unknown SIP-signalling errors that were caused by other carriers. Future work will comprise an investigation of these SIP-signalling errors, since they appear to be potential predictors for telephony fraud. This especially concerns so-called inter-carrier fraud.

Still, the developed system needs more testing. It appears that the thresholds have to be specified more precisely. As these values rely on experiences, the software has to be run in a productive environment with near real-time data to exactly determine the thresholds, in order to increase the detection probability. The final decision, if the results detected by the system are fraud, still relies on a human operator judging each case. Much harm could be done by automatically blocking innocent customers due to false positive classification results. With the presented approach, our system is able to conduct most of the analysis necessary to detect fraud by itself. Therefore, the probability that the delivered results indicate real fraud cases is already high. Given the modular implementation, the system can be easily extended. More rules, that is to say more filters, can be integrated with no effort. The more distinct the filters are that analyse the incoming data, the more likely it is to detect fraud before too much damage is done.

Granted that the presented system is tested more thoroughly, it will be capable to be used on a Next Generation Network for high-performance fraud detection. Its application will possibly improve the detection of telephony fraud, and it is worth considering for use by telecommunication service providers.

It should be stressed that the solution is specifically tailored to the needs of the collaborating service provider. From their perspective, the presented approach represents a major achievement concerning fraud detection in their practice, compared to the previously used solution. In comparison to the related work that has been presented in Section 2, our approach is especially distinguished by its ease of use in relation to the obtainable detection rate. The latter is already sufficient for this special case of application. For the expert personnel that actually use our system, defining or adjusting the rules is intuitive and relatively straight-forward to do. In this regard, our solution differs from other, more general applications of rule-based approaches. Neural networks, on the other hand, are often regarded as non-intuitive, because the inner structure of a trained neural network is not easily interpretable (Hastie, Tibshirani and Friedman, 2008). Hence, adjusting such a system could be harder for the intended users to do. On the other hand, future work will

especially comprise the integration of other methods from the field of machine-based learning into our solution, while keeping the focus at user-friendliness.

6. Acknowledgment

This work has been performed for the “Fraud Detection” project of Hochschule Darmstadt - University of Applied Sciences. The project is partially funded by the “Bundesministerium für Bildung und Forschung” (BMBF) and supported by the “Center for Advanced Security Research Darmstadt” (CASED). The authors additionally would like to acknowledge the support of toplink GmbH, which made this work possible.

7. References

- Bundesministerium der Justiz (2009), “Bundesdatenschutzgesetz“ in der Fassung vom 14. Januar 2003, zuletzt geändert am 14. August 2009,“ Berlin.
- K. C. Cox, S. G. Eick, G. J. Wills and R. J. Brachman (1997), “Visual data mining: Recognizing telephone calling fraud,” in: *Data Mining and Knowledge Discovery*, vol. 1, no. 2, pp. 225–231.
- Communications Fraud Control Association (2009), “2009 global fraud loss survey,” <http://www.cfca.org/>, (Accessed 01 September 2011)
- T. Hastie, R. Tibshirani and J. Friedman (2008), *The elements of statistical learning*, 2nd edition, Springer, Berlin/Heidelberg.
- J. Hollmén and V. Tresp (1999), “Call-based fraud detection in mobile communication networks using a hierarchical regime-switching model,” in: *Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems 11 (NIPS 1999)*. Morgan Kaufmann, 1999; pp. 889–895.
- Y. Kou, C.-T. Lu, S. Sirwongwattana and Y.-P. Huang (2004), “Survey of fraud detection techniques,” in: *Proceedings of the 2004 IEEE International Conference on Networking, Sensing and Control (ICNSC 2004)*. IEEE, 2004; pp. 749–754.
- D. Olszewski (2011), “Fraud detection in telecommunications using Kullback-Leibler divergence and latent Dirichlet allocation,” in: *Proceedings of the 10th International Conference on Adaptive and Natural Computing Algorithms (ICANNGA 2011)*. Springer, 2011; pp. 71–80.
- T. Padmaja, N. Dhulipalla, R. S. Bapi, and P. R. Krishna (2007), “Unbalanced data classification using extreme outlier elimination and sampling techniques for fraud detection,” in: *Proceedings of the 15th International Conference on Advanced Computing and Communications (ADCOM 2007)*. IEEE Computer Society, 2007; pp. 511–516.
- P. S. Foundation (2011), “Python programming language - official website”, <http://www.python.org>, 1990-2011, (Accessed 12. 12. 2011)
- S. Qayyum, S. Mansoor, A. Khalid, K. Khushbakht, Z. Halim and A. Baig (2010), “Fraudulent call detection for mobile networks,” in: *Proceedings of the 2010 International Conference on Information and Emerging Technologies (ICIET 2010)*. IEEE, 2010.
- S. Rosset, U. Murad, E. Neumann, Y. Idan and G. Pinkas (1999), “Discovery of fraud rules for telecommunications challenges and solutions,” in: *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 1999)*. ACM, 1999; pp. 409–413.
- I. Ruiz-Agundez, Y. Peña and P. Garcia Bringas (2010), “Fraud detection for voice over ip services on next-generation networks,” in: *Proceedings of the 4th Workshop in Information Security Theory and Practice (WISTP 2010)*. Springer, 2010; pp. 199–212.
- TELES (2006), “Teles.icdr, S48-S2000 series,” Teles Communication Systems, 2006.
- M. Taniguchi, M. Haft, J. Hollmén and V. Tresp (1998), “Fraud detection in communication networks using neural and probabilistic methods,” in: *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 1998)*. IEEE, 1998; pp. 1241–1244.
- S. Tindal (2009), “VoIP hackers strike perth business,” ZDNet, Jan. 2009. <http://www.zdnet.com.au/voip-hackers-strikeperth-business-339294515.htm>, (Accessed 05 August 2011)
- Zar, J. et al. (2005), “VOIPSA - VoIP security and privacy threat taxonomy, public release 1.0”, <http://www.voipsa.org/activities/taxonomy.php>, October 2005, (Accessed 05 August 2011)

A Criteria-Driven Method for Architecting Domain-Specific IE Applications

T. Neef, S. Morana, B. G. Humm

Department of Computer Science
Hochschule Darmstadt – University of Applied Sciences, Darmstadt, Germany
{tobias.neef, stefan.morana}@stud.h-da.de, bernhard.humm@h-da.de

Abstract

This paper proposes a method for architecting domain-specific information extraction (IE) applications focusing on a good cost/benefit ratio for a concrete domain. The method uses criteria to recommend the appropriate use of rule-based or machine learning based methods in the IE application architecture. By using an example from the tourism domain, the paper describes how the evaluation criteria can be applied in practice. An evaluation of the costs and benefits indicates a good IE recognition rate with reasonable development effort.

Keywords

NLP, supervised Information Extraction, tourism, architecture

1. Introduction

Companies have increasing need to access semantic information contained in natural language text. This is an ongoing trend in the last decade which is mainly motivated by the fact that the Internet enables companies to access massive amounts of information available in textual form. Semantic information in the context of this paper is knowledge contained in documents which are relevant to users of a specific domain. *Natural Language Processing (NLP)*, and in particular *Information Extraction (IE)* is the technique to extract semantic information from natural language text. A profit-oriented institution that develops an IE application needs to define a domain-specific architecture with a good cost/ benefit ratio.

Most IE publications, today, focus on approaches which solely increase the recognition rate, i.e., the benefit of IE applications. However, the development costs induced by those approaches are rarely considered. In this paper, we define a criteria-driven method for specifying the architecture of domain-specific IE applications with a well-balanced cost / benefit ratio.

The paper is structured as follows. Section 2 presents related work including state-of-the-art domain-specific IE approaches. Based on these approaches, the method for architecting domain-specific IE applications is defined (Section 3) and then applied to a specific domain (Section 4). The evaluation (Section 4.3) shows the development effort (cost) and recognition rate (benefit) of the resulting application. Section 5 concludes the paper.

2. Related Work

Recent work has shown the capabilities of various approaches for IE in domain-specific (supervised) scenarios. Those approaches can be categorized into two logical components, *Entity Mention Detection (EMD)* and the *Relation Mention Detection (RMD)* (Surdeanu et al., 2011). EMD groups the techniques used for detecting (named) entities. RMD detects (semantic) relations which connect entities.

Publications like (Surdeanu et al., 2008) and others focus on applications which are solely based on *Machine Learning (ML)*. In ML, the extraction of information is based on an ML model which was trained with sample data, a so-called *corpus*.

IE systems like ANNIE (Isabelle et al., 2001) do not use machine learning. Instead, the system relies on domain-specific resources like gazetteers or taxonomies for EMD. ANNIE is capable of doing morphologic normalization and basic rule-based co-reference resolution in order to increase the entity recognition rate. Extended versions of ANNIE use the annotation pattern language JAPE which is a finite state transducer (Cunningham et al., 2000). With JAPE, rules can be defined which use syntactic information to realize rule-based EMD and RMD. (Wyner and Wim, 2011) shows how this

approach is applied to the legal domain. They propose a linguistically motivated system which is based on Phrase Structure Parses to represent syntax.

(Surdeanu et al., 2011) argue that in supervised scenarios, the performance of a domain-specific IE application can be optimized using domain-specific components like rules and gazetteers.

In summary, both ML-based and rule-based approaches as well as mixed approaches have been applied to domain-specific IE. The focus of research in those areas was on the improvement of recognition rate measures. The following section proposes evaluation criteria which also consider the effort spent on learning and customizing a system.

3. A Criteria-Driven Method for Architecting Domain-Specific IE Applications

3.1. Overview

The method takes as input the prioritization of certain criteria and influencing factors of the problem domain. The outputs are architecture recommendations for the IE application to be constructed. See Figure 1.

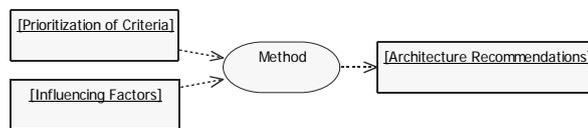


Figure 1: Inputs and outputs of the method

Criteria, influencing factors and architecture alternatives are described in the following sections.

3.2. Criteria

General criteria for architecting applications are *costs* and *benefits* where a good cost / benefit ratio is aspired. In the context of IE applications, these criteria can be refined as follows.

1. *Recognition rate*: a high recognition rate is the main benefit of an IE application
2. *Effort*: the development effort is the main cost factor for an IE application and should be as low as possible. It can be split up into two factors.
 - 2.1. *Customization effort*: IE applications are usually built on top of off-the-shelf IE components that need to be customized. The classic programming effort is usually relatively low.
 - 2.2. *Learning effort*: The use of off-the-shelf IE components is usually complex and requires proficiency in NLP and the technique used. Depending on the NLP expertise in the development team, the learning effort may be substantial.

In this paper, we do not consider the cost factor for software licences since our method is independent of concrete off-the-shelf products and their pricing models.

Depending on the domain under consideration, either criterion may be prioritized differently. In one project, the recognition rate has top priority and high costs may be acceptable. In other projects, a reasonable recognition rate is acceptable but the costs must be limited. The *prioritization of the criteria* is an important input for architectural decisions.

3.3. Influencing Factors

Apart from the prioritization of the criteria, there are other factors influencing architectural decisions for IE applications.

1. *Linguistic complexity*: the domain under consideration may involve different linguistic complexities. Aspects are, e.g., grammatical correctness incl. the occurrence of typing errors, the use of domain-specific terminology, the writing style (length of sentences, nesting depth), or the occurrence of co-references.
2. *Team expertise*: The expertise of the development team regarding NLP and IE technologies and concrete off-the-shelf components has a strong influence on the learning curve and the development effort.
3. *Domain-specific resources*: The availability of domain-specific resources like, e.g., dictionaries (taxonomies, ontologies) or corpora has an influence on IE approaches to be used.

3.4. Reference Architecture for IE Applications

Figure 2 shows a reference architecture for IE applications, i.e., a blue print that shows essential components, inputs and outputs of IE applications in general.

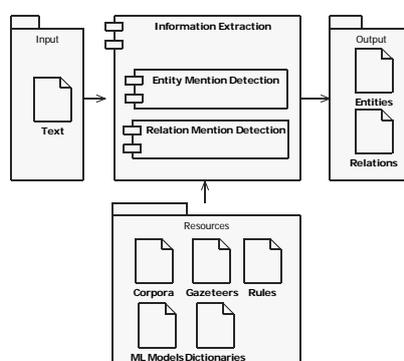


Figure 2: Reference architecture for IE applications

The main *input* for an IE application is *text* in a natural language. The main *outputs* are detected *entities* and *relations*. The main components of an IE application refer to the main *IE tasks*: *Entity Mention Detection (EMD)* and *Relation Mention Detection (RMD)*. *Resources* may be used to configure the IE components: *corpora*, *ML models*, *gazeteers*, *dictionaries*, and *rule sets*.

EMD and RMD may be performed by both, *machine learning (ML)* and *rule-based (RB)* approaches each using different resources. See Table 1.

IE Approaches IE Tasks	Machine Learning (ML)	Rule-Based (RB)
Entity Mention Detection (EMD)	Corpora, ML Models (Gazeteers, Dictionaries)	Rules, Gazeteers, Dictionaries
Relation Mention Detection (RMD)	Corpora, ML Models	Rules

Table 1: Resources used by IE approaches for IE tasks

3.5. Architecture Recommendations

Our method gives architecture recommendations for IE applications depending on the prioritization of criteria and influencing factors of the domain under consideration. The following Table 2 gives an overview.

Topic	Input: Prioritization of criteria and influencing factors	Architecture Recommendation			
		EMD		RMD	
		ML	RB	ML	RB
(A.) Prioritization of criteria	(A.1) Top priority on recognition rate	✓		✓	
(B.) Linguistic complexity	(B.1) High linguistic complexity	✓		✓	
	(B.2) Low linguistic complexity		✓		✓
(C.) NLP expertise	(C.1) Limited NLP expertise in development team		✓	✗	✓
	(C.2) High ML experience in development team	✓		✓	
	(C.3) High linguistic experience in development team		✓		✓
(D.) Domain- specific resources	(D.1) Dictionary available	✓	✓		
	(D.2) Annotated corpus available	✓		✓	

Table 2: Overview of architecture recommendations

The column “Input” describes the prioritization of criteria and influencing factors in a particular application domain. The column “Architecture Recommendation” indicates a preferable selection of an IE approach (machine learning or rule-based) for a particular IE task (entity mention detection or relation mention detection). A tick (✓) denotes a suitable approach, a cross (✗) denotes an approach which is to be avoided. If, in a concrete application domain, different recommendations are in conflict then the application architect has to make an informed decision. The following paragraphs will help the architect in making such a decision.

Basics

Named entity recognition (NER) based on gazetteers can be broken down into two segments. The first one includes pre-defined lists of categorized words (*gazetteers*). If there is no doubt to which category those lookups belong to then they can be marked as Named Entities. Often, it depends on the context of a token to which category the Named Entity can be assigned. Also problems like word sense ambiguity, incorrect spelling or co-references make it unfeasible to solely rely on gazetteer lists. Therefore, gazetteer-based systems like ANNIE include context-specific rules in order to yield better EMD performance.

Another approach to EMD is the use of machine learning (ML) based NER as pioneered by (Lafferty, 2001). As with all ML approaches, a vast amount of pre-annotated documents of the given domain have to be available to yield good results. But those systems can be significantly more robust against various forms of misspellings. Additionally, they can learn the correct categorization for the given context and they already solve some co-reference related problems.

When evaluating the complexity of both approaches it soon becomes obvious that the complexity characteristics are not linear. Instead they are a function based on the quality of input data which is available for a given domain, and on the linguistic complexity of the domain.

Linguistic Complexity (B.)

The rules in a RB approach have to be created manually. The customization effort for the system increases with the linguistic complexity of the domain documents. A ML-based application uses a corpus. The complexity of annotating a corpus is not related to the linguistic complexity of the document. As a consequence, we recommend ML-based approaches for EMD and RMD when the linguistic complexity is high (B.1).

Prioritization of criteria (A.)

Rule-based and ML-based approaches can also be combined. According to (Surdeanu et al., 2011), the combination of gazetteers and ML-based NER is required to optimize the recognition rate of a domain-specific IE application. However, this combination has an effect on the system complexity and, hence, on learning and customization effort. If the recognition rate is prioritized (A.1) then the combination of ML-based and RB approaches is recommended.

Domain-specific resources (D.)

In the gazetteer approach, gazetteer may be derived from resources like dictionaries or taxonomies. If such resources do not exist then domain-specific gazetteer lists need to be constructed manually by domain experts. Depending on the domain and the number of relevant entities, this may lead to a customization effort which is not acceptable anymore (Kozareva, 2006) (D.1).

The initial customization effort of the ML-based NER is mainly influenced by the effort required to annotate documents. They are the input for the learning algorithms. If those annotated documents are not available then the effort of creating a sufficiently large corpus of good quality is a high initial investment.

When looking at the customization effort of both approaches we argue that gazetteer-based applications have a relatively low initial customization effort. It increases if the quality of the domain specific dictionaries is low or no dictionaries are available. The initial customization effort of ML based methods is higher but does decrease when an annotated, similar corpus is already available (D.2).

NLP Expertise (C.)

The learning effort of the gazetteer-based approach is mainly related to the linguistic complexity of the domain and the annotation features available. The learning effort of creating gazetteer lists is minimal as long as the creator of the lists has enough domain knowledge. In contrast, the EMD related rules require a solid understanding of rule systems like JAPE (Cunningham, et al., 2000), as well as the features which are used as input for the rule system. Features can be part-of-speech (POS) information as well as various representations of syntax.

The learning effort of ML-based NER is mainly influenced by the abstractions the tools offer to hide the complex ML algorithms from the developer. This has been a problem which was actively worked on since research started to focus on this area (Lafferty, 2001). Therefore, tools have emerged on the market like the Stanford NER (Finkel, et al., 2005) or GATE's Batch Learning PR (Li, et al., 2005) which are useable for an average developer to create a domain-specific NER models.

An important factor in the learning effort of the EMD component is the background of the developers. If they have a more linguistic background the gazetteer based approach is recommendable (C.3). When the users have some background in statistics and machine learning, they can soon get productive with the ML-based approach (C.2). Both techniques have a certain level of initial complexity which needs to get resolved.

In ML-based RMD, features like Named Entities are required which are detected by preprocessing steps or EMD. Those features and a corpus annotated with relations is the input for a learning algorithm. The combination of these features and the customization is an expert task and, from our perspective, no tool has gained enough traction to be used outside a specific research group (C.1).

Rule-based RMD is mostly based on syntax information. This information is then processed by a rule system in order to find the semantic relationships. There are two commonly used syntactic representations: *phrase structure parses* and *dependency grammars*. Phrase structure parses are trees that are optimized to represent the syntax in a detailed and linguistically correct way. Therefore, it has been used in many systems in the past. Publications like (de Marneffe and Manning, 2008) showed that for most non-linguists the dependency grammar representation is more appealing. They also proposed a dependency representation which is mainly motivated to provide semantically helpful information instead of showing all the syntactic details. Our experiments indicate that the acceptance of the dependency-based syntax representation is higher than the phrase structure parses for developers without a linguistic background. Further evaluation criteria of syntactic parsers and their representation have been developed by (Miyao et al., 2008).

Although the same customization effort tradeoffs apply for ML-based RMD as they do for EMD, we argue that, due to the lack of tool support in the ML-based RMD area, the learning effort and is too high to recommend the approach.

4. The Method Applied in a Sample Domain

4.1. The Hotel Domain

In order to demonstrate the proposed method, we introduce an IE problem in the tourism domain. This IE task was accomplished within the research project "Ontology-Based Text Mining" (OBTM) at Hochschule Darmstadt – University of Applied Sciences for the company HRS – Hotel Reservation Service, a leading hotel portal provider in

Germany. The task was to extract information about a hotel, its properties, and also about its rooms and their equipment. Thousands of textual hotel descriptions are available in catalogues or online web pages. HRS plans to use this information for offering their customers better search capabilities. In order to achieve this goal, a semantic understanding of natural language text is required. The OTA code list (OpenTravel Alliance, 2010), which is an existing classification system for the tourism domain, defines categories of hotel-related entries. All terms in the OTA code list have an identifier.

Code	Name	Category
GRI42	Room	Accommodation Profile, Accommodation Unit
HAC79	Sauna	Hotel Facility, Unit Facility

Table 3: Examples of OTA codes and categories

Hotel codes are assigned to categories depending on their roles. For example, the code “GRI42/Rooms” may be used to indicate the total amount of rooms. In the hotel description, this fact may be formulated as follows.

“Our hotel has 120 well-equipped rooms.”

In this case, it is expected to find a relation between the code “Room” and the number 120 which quantifies it.

RoomAmountRelation(GRI42/Room, 120)

Another role of a room is an Accommodation Unit which includes so called “Unit Facilities”.

“All our rooms contain a sauna for your personal pleasure.”

In this case it is expected to find relationships between the room and the units which are contained in this room.

RoomUnitRelation(GRI42/Room, HAC79/Sauna)

Finally, an entity may also be in the category “Hotel Facility”. In this case, the unit is not assigned to a room type. Instead, it is a unit of the Hotel:

“Concerning relaxation, you have the possibility to bring harmony between your body and your soul, thanks to the swimming pool, the Finnish sauna, the solarium, the fitness-room...”

While the fitness-room is quite obviously assigned to the hotel, it depends on the context to which entity the sauna is assigned. In this case, the sauna is related to the hotel.

HotelToUnitRelation(HAC79/Sauna)

But as seen in this section the sauna can also be assigned to a room.

This domain has a sufficiently high complexity to motivate the architecture which is shown in the following section.

4.2. Applying the Method to the Hotel Domain

The project members were postgraduate students (M.Sc.) without prior NLP experience and without much knowledge about either machine learning or linguistics (Criterion C.1). The focus of the project was to find most of the relations while keeping the costs low. As shown in the previous section, a domain-specific dictionary is available with the OTA code list (D.1). The linguistic complexity of the project is also an important factor for choosing the appropriate architecture. In the given domain, typing and grammar errors are rare. Additionally, co-references could appear theoretically but could not be identified in the test corpus. Because most hotel descriptions are created by the hotel marketing departments, the sentences tend to be lengthy and sometimes have a high nesting depth. Overall, the linguistic complexity is considered relatively low (B.2).

Entity Mention Detection: Because the OTA code list was available as a mature domain dictionary, it was easy to create gazetteers (D.1). Including the gazetteers is also useful because the team had no prior NLP experience. This makes gazetteer lists a good starting point for EMD (C.1). The only customization effort related to the gazetteers was the consideration of synonyms. This was done manually based on sample data. The hotel domain is not linguistically complex. Therefore, it is not required to include a ML-based NER component (B.2).

Relation Mention Detection: The selection of the RMD approaches is also related to the considerations in EMD. No criteria would motivate a ML-based RMD system with the given project parameters.

According to the recommendations from the method, the application architecture chosen included rule-based RMD and EMD together with gazetteers.

4.3. Evaluation

An evaluation in terms of recognition accuracy was performed. The corpus contained a total of 227 hotel descriptions for 124 hotels. The corpus was split into two disjoint parts. One part containing 100 documents was used for optimizing the customization of the application, e.g., adding new synonyms to the gazetteer list and creating syntax patterns. In order to test the effect on new documents, the recognition rate accuracy was calculated using the second part of the corpus. For assessing the results, we use *Precision (P)*, *Recall (R)* and *F-Measure (F)*, which are commonly applied measures from Information Retrieval. *P* describes the correctness, *R* describes the completeness and *F* is their weighted harmonic mean. Table 5 shows the results.

	F (EMD)	F (RMD)
Gazetteers based on OTA	0.75	0.72
Gazetteers based on OTA and synonyms	0.87	0.85

Table 4: Evaluation result

Generally, the results are satisfactory for the hotel domain. The data shows that the RMD recognition rate is mainly capped by the EMD recognition rate. By enhancing the gazetteers with synonyms of the OTA codes, the F-measure could be improved by 0.12 (EMD) and 0.13 (RMD), respectively.

The customization was implemented by four postgraduate computer science students (M.Sc.) without prior NLP experience in a six month timeframe. This indicates that the complexity of this domain could also be handled by a development organization in an industry project.

5. Conclusions

This paper presents a criteria-driven method for architecting domain-specific IE applications. We have developed evaluation criteria which balance various IE techniques in terms of costs and benefits. To our best knowledge, such a method that takes into account IE application development costs has not been presented before. We showed that this approach has been valuable for a real-world application scenario from the tourism domain. We argue that other domains could also benefit from it.

As future work, we plan to apply this method in more IE projects in order to review the validity of our statements. We assume that the tooling for ML-based RMD will emerge which would require a re-evaluation of the presented criteria.

6. Acknowledgement

This work was funded by HRS – Hotel Reservation Service.

7. References

- Cunningham, H., Maynard, D., and Tablan., V. (2000) “JAPE: a Java Annotation Patterns Engine (Second Edition)”, Technical report CS-00-10, 2000, University of Sheffield - Department of Computer Science, Sheffield.
- Finkel, J. R., Grenager, T., and Manning, Ch. (2005) “Incorporating non-local information into information extraction systems by gibbs sampling”, *Proceedings of the 43rd Annual Meeting of the Association*, 2005, Association for Computational Linguistics, Ann Arbor, Michigan, pp363-370.
- Isabelle, P., Cunningham, H., Maynard, D., Bontcheva, K., and Tablan, V. (2001) “GATE”, *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics 2001*, Association for Computational Linguistics, Stroudsburg, PA, USA, pp168-175.
- Kozareva, Z. (2006) “Bootstrapping named entity recognition with automatically generated gazetteer lists”, *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop*, Association for Computational Linguistics, 2006 Stroudsburg, PA, USA, pp15-21.

- Lafferty, J. (2006) "Conditional random fields: Probabilistic models for segmenting and labeling sequence data", Proceedings of the 18th International Conf. on Machine Learning, 2001, Morgan Kaufmann, San Francisco, CA, pp282-289.
- Li, Y., Bontcheva, K., and Cunningham, H. (2005) "SVM Based Learning System For Information Extraction", Proceedings of Sheffield Machine Learning Workshop, LNCS, 2005, Springer Verlag, Heidelberg, pp319-339.
- Marneffe, M., and Manning, Chr. D. (2008) "The Stanford typed dependencies representation", Coling 2008: proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation, 2008, Association for Computational Linguistics, Stroudsburg, PA, USA, pp1-8
- Miyao, Y., Sætre, R., Sagae, K., Matsuzaki, T. and Tsujii, J. (2008) "Task-oriented Evaluation of Syntactic Parsers and Their Representations", Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, The Association for Computer Linguistics 2008, Columbus, Ohio, USA. pp46-54
- OpenTravel Alliance (2010) OpenTravel Implementation Guide: Executive Summary, No. 1.5, 2010. http://www.opentravel.org/Resources/Uploads/PDF/OpenTravel_ImplementationGuide_v1.5_ExecSum.pdf, (Accessed 2 November 2011)
- Surdeanu, M., Johansson, R., Meyers, A., Màrquez, L., and Nivre, J. (2008) "The CoNLL-2008 shared task on joint parsing of syntactic and semantic dependencies", Proceedings of the Twelfth Conference on Computational Natural Language Learning. 2008. Association for Computational Linguistics. Stroudsburg, PA, USA, pp159-177
- Surdeanu, M., McClosky, D., Smith, M. R., Gusev, A., and Manning, C. D. (2011) "Customizing an Information Extraction System to a New Domain", Proceedings of the Workshop on Relational Models of Semantics, 2011, Association for Computational Linguistics. Stroudsburg, PA, USA.
- Wyner, A., and Wim, P. (2011) "On Rule Extraction from Regulations", Proceedings of the 24th International Conference on Legal Knowledge and Information Systems. IOS Press, Amsterdam, 2011.

Attacks to ZigBee and Wireless Sensor Networks Honeypots for Detection and Response

J. Markert¹, M. Massoth², K.-P. Fischer-Hellmann², S. M. Furnell¹

¹Plymouth University, Plymouth, United Kingdom

²Hochschule Darmstadt – University of Applied Sciences, Darmstadt, Germany

e-mail : jurgen.markert@plymouth.ac.uk

Abstract

ZigBee is a new communication technology and offers opportunities not only for developers, but also emerges as a possible risk containing flaws for an attacker to focus on. There is a trend in this field to develop methods for tracking intrusions and unauthorized access by an attacker, and for detecting attempted attacks against the stability of wireless sensor networks with intrusion detection systems. These systems represent the current state of the art in the field of research on detecting attacks on wireless sensor networks. We therefore propose implementing honeypots for ZigBee networks, which would be an alternative approach to intrusion detection systems, and which would prove to be ideal for the development of appropriate countermeasures. We feel that not enough research has been done so far in the development of ZigBee honeypots to offer opportunities for specific analysis and intercepting attacks.

Keywords

Wireless Sensor Network, ZigBee, Intrusion Detection System, Honeypot

1. Introduction

The next industrial revolution will be the internet of things, machine to machine communication, “smart homes”, “smart buildings” or even “smart cities” (Sorensen 2010). The concept is deceptively simple: Everything can be connected to anything else. Anything provides services and can be monitored and controlled. This evolution is already laid out and well on its way. IPv6 and other technical prerequisites are already available to cope with the challenge of having more and more systems to be addressed in a single network.

Yet, these networks don't necessarily have to be wire-bound. A lot of networking technologies have their focus on wireless communication channels. Many companies have worked together in the development of Bluetooth as a wireless standard. ZigBee was likewise introduced and developed to be an ideal standard for wireless sensor and control networks (ZigBee Alliance 2008). ZigBee is an additional feature set to the IEEE 802.15.4 standard and defines additional layers on top of it (IEEE 2006).

IEEE 802.15.4 defines physical radio and a MAC layer, providing a simple packet data protocol for lightweight wireless networks. ZigBee adds the layers for logical network, security and application software. ZigBee determines the API, and the ZigBee Alliance certifies ZigBee-compatible devices guaranteeing their interoperability (Elahi 2010).

New ways of interaction become available through ZigBee, offering great advantages and a lot of potential to save energy and make life much more comfortable, but also requires that underlying structures be robust, reliable, safe, and secure. The key aspects of network technologies are commonly the same: application developers are expected to protect communications in order to attain the classic information security requirements: confidentiality, integrity, and availability.

2. Motivation

The analysis of the security and reliability of ZigBee networks has shown that these aspects have not yet been resolved satisfactorily, depending on the chosen implementation. But not all of the above requirements are present in the various configurations. The background to this is as easy as might be: the intended purposes of wireless sensor networks vary widely. One of them is low power consumption, running independently for the longest possible time, another one is for these devices to be as interoperable, compatible and easy to use as possible. This is in direct conflict of nearly every security feature, because security needs computational power and thus results in a higher energy consumption. These

features also limit their ease of use and render interoperability problematic: Encryption details have to be distributed and managed, requiring a bigger feature set, more network traffic.

The paper by Fabbriatore et al. (2011) describes these problems in the area of smart homes. The authors see several problems for existing home automation solutions involving ZigBee. They propose using additional cryptographic extensions on top of ZigBee, or to change the protocol, should ZigBee remain insecure for their purposes.

3. Attack categories

Common attacks on wireless sensor networks (WSN) as described in current publications and literature are part of a range of working and well-known attacks on ZigBee networks: Eavesdropping, replay attack, sinkhole attack, selective forwarding, network flooding, firmware modification and code injection (DePetrillo 2009), (Goodspeed 2007), (Gu & Noorani 2008), (Masica 2007), (Wright 2009), (Yang et al. 2008). These attacks can be subdivided into three categories:

3.1. Resource drain attacks

The first category shall be termed “resource drain attacks”. These attacks target the power of network devices as shown in figure 1 (replay-attack). In a battery-driven system, the main goal of an attacker is to exploit the system's power consumption against the system itself. An attacker might render wireless sensor network unusable by making parts of it run out of battery power through the modification of networking schematics, with the propagation of new routes, and by incessantly sending and requesting data. Batteries are a limited resource, network capacity is likewise limited. An increasing network load limits cripples the already limited throughput of the network.

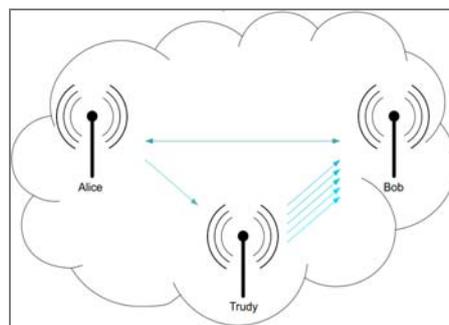


Figure 1: Power drain through replay attack

The authors Raj and Thilagavithy (2012) propose a solution to a specific type of network load caused by so called jamming attacks in wireless sensor networks. Due to the amount of commonly available radio channels in these networks, shifting to another network channel is suggested each time a jamming attack is detected. The authors claim that a residual network activity should keep the channel busy for a limited time during the short interval in which the jamming is not taking place as to make the jamming attacker believe that the jamming attack is successfully disturbing the WSN communication. This behaviour could also be described as deception, or, simply put, a honeypot or honeynet. This could even also attract an attacker and act as a decoy for the attacker.

3.2. Network attacks

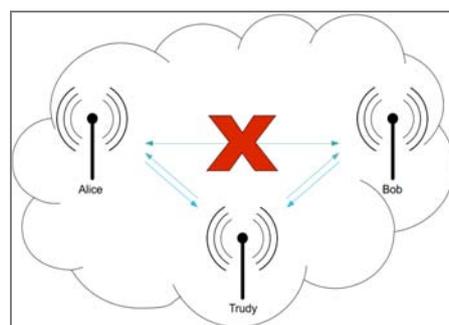


Figure 2: Network redirection for man in the middle scenario

The second category is called “network attacks”, describing the attacker's goal of modifying the network's routing structure. By this, the attacker aims to gain valuable information through redirection. He becomes therefore a “man in the middle” listening to any traffic.

3.3. Hardware-specific attacks

The third category is called “hardware-specific attacks”, standing for the goal to bring the networking nodes under complete control of the attacker (e.g. firmware modification as shown in fig. 3). This could be achieved with physical attacks like attaching wires in order to extract encryption keys from a network node. Attacks on hardware can also be an attack on firmware, resulting to firmware-modification. This is possible through security flaws in software-components of the nodes.

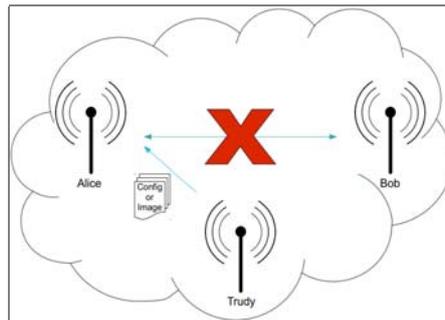


Figure 3: Firmware-modification attack

All of the three attack types mentioned above already exist. There is even a “proof of concept” worm that exploits network node communications and propagates itself over the network.

Goodspeed et al. (2012) have presented new results in the area of WSN hacking. They improved the killerbee API and wrote extensions to execute “wardriving” scenarios on ZigBee networks. They extended existing tools until all 16 ZigBEE channels could be monitored in parallel. They then combined this with a GPS signal logger for “wardriving” around the campus of Dartmouth. They also wrote semi-automatic tools to attack the networks with various different attack types. This development should make abundantly clear, that the security threats for ZigBee are real and serious and it is the conclusive proof that solutions to detect the attacks is are of crucial necessity.

4. Countermeasures by design

Several security features are implemented in ZigBee by design. “ZigBee PRO 2007” defines security mechanisms for authentication, encryption and trust centre functions (Elahi 2010).

Research on these features has however revealed some shortcomings. Security features might be circumvented or their imperfection could be used against the system. Successful attacks might last as long as the attacker wants. These can even not be analysed, because usually, there is no log of network activities and hence, attackers will leave no traces. There is no “packet capture” of attacks, and the probed exploit code is not available in the wake of an intrusion attempt. An attacker will not get noticed and therefore, will remain unidentified. It is hardly possible to gather the leftover information for forensic analysis.

There are clearly still problems in the process of improving the security features of ZigBee networks. Having blind faith and trust in the existing countermeasures is greatly unjustified. Radmand et al. (2010) showed in their comprehensive paper, that the cryptography used in ZigBee is not encompassing enough to cover all possible attacks. Some of the problems like confidentiality are solved encrypting the data stream. Yet, problems on the network layer such as replay attacks still remain. They also state, that the manufacturers should provide a minimum of security, due to problems with encryption keys stored in plain-text in the nodes. These could be extracted by an attacker tampering with physically captured ZigBee network nodes. There is hence a need for an intrusion detection system, and also for system reporting break in attempts.

5. Intrusion Detection System

Intrusion attempts are reported in classical network security solutions by so-called intrusion detection systems (IDS). They use only passive inspection to monitor all network activities for violation of regular use, and have no other function in the net. Intrusion detection systems rely on predefined policies describing normal network activity and normal behaviour. An intrusion detection system is generally a single point in the network, but it might additionally also be distributed on several hosts of the network.

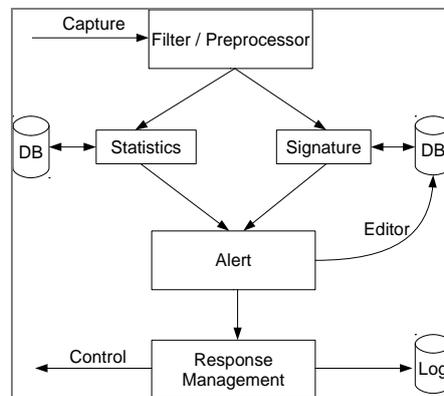


Figure 4: Intrusion detection system (IDS) working scheme

It might also stop intrusion or filter or modify malicious traffic as an active part of the network and is in this case most often called an intrusion prevention system (IPS). In an IPS, the controlling aspect of figure 4 is much more elaborate than in an IDS. All of these setups provide logging and reporting. In an IDS, every data packet has to be regarded and analysed. Its purpose is to spot the packets of an attack within a large amount of regular and normal network traffic. Processing every packet, its inspection, and the detection of attacks by matching patterns all require a lot of resources. In the context of ZigBee networks, and therefore battery driven network nodes with very limited resources, this is a clear drawback of an intrusion detection system approach in wireless sensor networks.

Kaplantzis et al. (2007) did a network simulation and had success in detecting selective forwarding attacks in wireless sensor networks by support vector machines (SVM), too. They also concluded that these attacks are the most difficult to accurately detect thus verifying assumptions made in previous research.

Sedjelmaci and Feham (2011) presented a novel IDS for the use in clustered WSNs. Using SVM, just like Kaplantzis et al. (2007) did, the essential point is to teach the detector nodes about normal network behaviour, such that they achieve a detection ratio of over 98% regarding abnormal network traffic.

Iwendi and Allen (2011) wrote a comprehensive paper on attacks on WSNs and demonstrated a way to simulate attacks with the intention of developing appropriate countermeasures. They propose in their conclusion the development of a security protocol for defensible WSN.

There is a definite need for security in wireless sensor network communications, and possible improvements should involve intrusion detection systems. The emerging outline can be briefly described as a sensor working on different network layers. Its purpose is to detect anomalies in the net and to report these attacks. This is supposed to provide a new technical instrument for reacting to this new attack scheme. This feature set forms the common core of an intrusion detection and prevention system, also called a honeypot.

6. Honeypots

A honeypot can be described with a single phrase: It is a trap. It is in general an active system in a network monitoring every connection to this part of the network. The results are written to log files and reported accordingly. The honeypot looks like a normal node of the network to every system passing by. To an attacker, however, it will appear as a valuable target.

The advantages of a honeypot might be summarized as follows: instead of inspecting and analysing every packet of the network like an IDS, and then deciding whether this has been an attack or not a honeypot simply regards every connection as a likely attack. This behaviour saves a lot of resources because not every packet needs to be processed,

inspected and filtered for attacks. Even in the context of ZigBee networks and therefore battery driven networks, the honeypot system could make do with very limited resources. One of the drawbacks of a honeypot system approach in wireless sensor networks could be that it exposes additional targets which might get corrupted and then be taken over by an intruder. But since any other part of the net could equally get compromised, this is a small drawback. The advantages of a honeypot far outweigh its disadvantages.

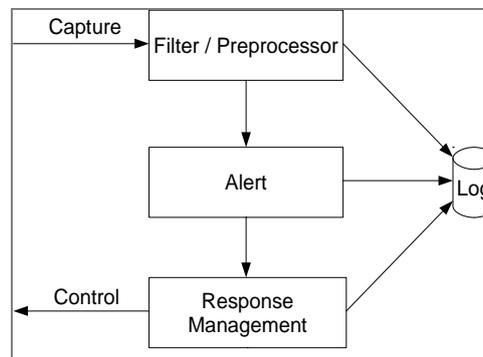


Figure 5: Honeypot working scheme

Prathapani et al. (2009) propose a honeypot system for the wireless mesh network standard 802.11s, then still in draft status. They set up the network simulator ns-2 and showed in their environment a very significant and successful result in network throughput with a high detection and low false positive ratio during a blackhole attack. Although these results were not proven to be correct in a real network, we assume that a similar result might be achieved in wireless sensor networks based on ZigBee standard communication.

Gupta et al. (2012) propose a honeypot technique for wireless mesh networks. A group of several honeypots are called a honeynet. The authors differentiate between different types of honeypots such as honeypots for research and honeypots for the productive environment. In detail they propose using a concept called honeyPHARM described by Hassan and Al Ali (2011). This solution has been specifically invented for the 802.11s draft. The authors Hassan and Al Ali (2011) propose a honeyPHARM, a distributed honeynet for the collection of malware in networks. Collected data will be used to discover new attack methods. This will be an addition to the nephentesPHARM system.

The authors Muraleedharan and Osadciw (2009) propose a framework using a honeypot technique combined with swarm intelligence for a battlefield monitoring application in a wireless sensor network. They state that traditional security schemes can not be applied in the field of WSN due to resource constraints. They had planned to build a simulation to prove this proposal, and validate the results.

The summary of all of these ideas leads to a proposed solution development of a honeypot for ZigBee networks without the drawbacks of a classical IDS. The concept of the honeypot will be presented below.

Several technical details are worthy of discussion. One of the proposed features of a honeypot should be that it might be temporarily part of the net at various different times. It should not always neighbor the same node in the network, because these nodes in the net might not be the target of the attacker.

One of the proposed features could also involve cloning. Those resources of special appeal to the intruder might duplicate themselves virtually during the course of an attack, expanding the target range and thus reducing the likelihood of an attack on the real network.

The system additionally features a feedback channel for reporting attacks. This can be probably done via a covert channel or other medium via a gateway. This might be done on other frequencies or channels like e.g. GSM, 5GHz WiFi, 868 MHz ZigBee, LAN and others.

A honeypot system should be a hardened system. Singh and Verma (2011) list a number of possibilities for hardening the WSN in the paper "Security For Wireless Sensor Network". They also show the list of unresolved problems, and state that these problems are difficult to solve.

A honeypot must not be recognizable as such to an attacker. An attacked honeypot should therefore react in the same way as any other attacked part of the wireless sensor network. This leads to the assumption, that its hardware capabilities should be the same. For safety and reliability reasons, it could be equipped with a larger battery or a wired power supply.

In their paper, Mostara and Navarra (2008) suggest assigning roles in wireless sensor networks with specific features for honeypots. They propose progressively changing the roles of the honeypot nodes over the lifetime of the network. This feature should be part of the WSN honeypot as well.

Regular clients are forbidden from using the honeypots for networking purposes, since connections to these honeypots would be regarded as an attack. Since the roles of nodes in this network should change over time between honeypot nodes and regular nodes, a predefined schedule will be used to guarantee a low false positive intrusion detection rate. This method renders the network harder to attack.

Detecting new attack methods is definitely possible with honeypots. The rest of the network will be left in a functional state while the honeypots are under attack. As an additional benefit, new countermeasures can be prepared while the attack is still ongoing. All these presented uses of honeypots are suitable to increase the reliability and availability of ZigBee networks.

Moreover, the recording unknown attacks will undoubtedly lead to the development of new countermeasures and hardening. The proposed honeypot concept for ZigBee networks combines new detection mechanisms offering new countermeasures for the stability and reliability of the network.

The installation and maintenance of a single honeypot or of a distributed honeypot network clearly requires additional effort. Yet, this effort should be made on top of common improvements to the wireless sensor network like configuration hardening and regular firmware updates. Some implementations offering over-the-air programming features for updating nodes already exist.

The detection of an attack by a honeypot and the insight in the behaviour of an intruder can be used to devise appropriate defence mechanisms or to generate updates for the complete network. Considering the publications over the last year discussing new attack schemes, there is an expected growing market for researchers to do research on security flaws simply due to the also growing opportunities for intruders to mess with the ZigBee networks (Cache et al. 2010). Honeypots will help in detecting and resisting these unavoidably upcoming attacks. It should be understood, though, that the security of a ZigBee network is not static, but rather an ongoing process to be kept up during its whole lifetime.

7. Conclusions

ZigBee offers some security features out of the box. Its weaknesses, however, have been explored only in part to the present day. It appears that only few people have been doing research in this area at all, and no one has ever applied honeypot methods in a ZigBee network. Our intended research will explore previously unconsidered and unregarded aspects of security threats against ZigBee networks and develop new defence mechanisms with the use of honeypots.

8. References

- Cache, J., Wright, J., Liu, V. (2010), "Hacking Exposed – Wireless", MC Graw Hill.
- DePetrillo, N. (2009), "Power Hungry People - Making Sense of New Critical Infrastructure Threats", http://proidea.maszyna.pl/CONFidence09/2/CONFidence2009_nick_de_petrillo.pdf.
- Elahi, A., Gschwender, A. (2010), "ZigBee Wireless Sensor and Control Network", Prentice Hall.
- Fabbricatore, C., Zucker, M., Ziganki, S. & Karduck, A. P. (2011), "Towards an Unified Architecture for Smart Home and Ambient Assisted Living Solutions".
- Goodspeed, T. (2007), "MSP430 buffer overflow exploit for wireless sensor nodes", <http://travisgoodspeed.blogspot.com>
- Goodspeed, T., Bratus, S., Melgares, R., Speers, R. & Smith, S. W. (2012), "Api-do: Tools for Exploring the Wireless Attack Surface in Smart Meters.", 2012 45th Hawaii International Conference on System Sciences., IEEE, pp. 2133–2140.
- Gu, Q., Noorani, R. (2008), "Towards self-propagate mal-packets in sensor networks", WiSec '08: Proceedings of the first ACM conference on Wireless network security, ACM
- Gupta, P., Rawat, P., Malik, S. & Gupta, S. (2012), "Securing WMN Using Honey pot Technique", (4), 235–238.
- Hassan, A. & Al Ali, M. (2011), "Collecting Malware From Distributed Honey pots - HoneyPHARM".
- IEEE, Institute of Electrical and Electronics Engineers - IEEE 802.15.4-2006 IEEE Standard (2006), <http://standards.ieee.org/getieee802/802.15.html>

- Iwendi, C. O. & Allen, A. R. (2011), "CIA Security Management for Wireless Sensor Network Nodes".
- Kaplantzis, S., Shilton, A., Mani, N. & Sekercioglu, Y. A. (2007), "Detecting Selective Forwarding Attacks in Wireless Sensor Networks Using Support Vector Machines."
- Masica, K. (2007), "Recommended Practices Guide For Securing ZigBee Wireless Networks in Process Control System Environments", <http://csrp.inl.gov/Documents>
- Mostarda, L., Navarra, A. (2008), "Distributed Intrusion Detection Systems for Enhancing Security in Mobile Wireless Sensor Networks", International Journal of Distributed Sensor Networks, Vol. 4, 2008.
- Muraleedharan, R. & Osadciw, L. A. (2009), "An intrusion detection framework for Sensor Networks using Honeypot and Swarm Intelligence.", Proceedings of the 6th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services. IEEE.
- Prathapani, A., Santhanam, L. & Agrawal, D. P. (2009) "Intelligent Honeypot Agent for Blackhole Attack Detection in Wireless Mesh Networks".
- Radmand, P., Domingo, M. & Singh, J. et al. (2010), "ZigBee/ZigBee PRO Security Assessment Based on Compromised Cryptographic Keys.", 2010 International Conference on P2P, Parallel, Grid, Cloud and Internet Computing., IEEE, pp. 465–470.
- Raj, J. S. S. & Thilagavathy D. (2012), "Security Threats And Jamming Attacks Of Multi Channel Wireless Sensor Networks.", International Journal of P2P Network Trends and Technology- Volume 2 Issue 1- 2012 (2), 27–31.
- Sedjelmaci, H. & Feham, M. (2011), "Novel Hybrid Intrusion Detection System For Clustered Wireless Sensor Network.", International Journal of Network Security & Its Applications 3 (4), 1–14.
- Singh, S. & Verma, H. K. (2011), "Security For Wireless Sensor Network.", International Journal of Network Security & Its Applications (3), 2393–2399.
- Sorensen, S. (2010), "The Sustainable Network", O'Reilly
- Yang, Y., Zhu, S., Cao, G. (2008), "Improving sensor network immunity under worm attacks: A software diversity approach", MobiHoc '08: Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing.
- Yüksel, E., Nielson, H. R., Nielson, F. (2010), "A Secure Key Establishment Protocol for ZigBee Wireless Sensor Networks", Oxford University Press.
- ZigBee Alliance (2008), "Latest ZigBee Specification Including the Pro Feature Set", <http://zigbee.org/Products/DownloadZigBeeTechnicalDocuments.aspx>, Std. 053 474r17

Development of Three-Dimensional User Interfaces based on Low-Cost Inertial Navigation Systems

M. Muentner¹, M. Haid¹, T. Chobtrong, E. Guenes¹, M. Kamil¹

¹ Competence Center for Applied Sensor Systems (ccass)
Hochschule Darmstadt – University of Applied Sciences, Darmstadt, Germany
e-mail: moritz.muentner@gmail.com

Abstract

The aim of this paper is to find a suitable technology that allows a three-dimensional position determination within a specific space using today's available sensor systems. Starting with a brainstorming, the key technologies inertial sensors, image processing and RFID were selected and elaborated in detailed concepts by methodological approach. Finally, the detailed concepts were validated on the basis of different application scenarios, as well as advantages and disadvantages discussed. The results revealed that the concept of an inertial navigation system has the most advantages in terms of feasibility, costs and computational effort. The camera system works precisely under ideal conditions but the system was not able to convince due to high processing power and issues of shadowing effects. The RFID concept offers potential for the future because the determination of position using RFID is a current subject of development. However, RFID was not able to prevail because of its low resolution and its liability to interferences. This paper sums up the key features of the inertial navigation system.

Keywords

INS, inertial navigation system, user interface, indoor navigation, object tracking

1. Introduction

In recent years, the interface between human and machine was only possible by means of external input devices in order to interactively work with a machine. Nowadays, screen and input peripherals are combined in a compact and mobile unit, the touch screen. Touch screens are a convenient tool for communication between human and machine. This enables the solution of two-dimensional communication tasks, such as the acquisition of objects whose main distinctive characteristics are located in the plane (e.g., hexagonal nuts, screws, circular). To recognize objects or to distinguish different levels of stacked objects, a touch screen requires the compulsory acquisition of the third dimension. A general three-dimensional object detection and representation therefore requires new technologies that are subject to this paper.

2. Inertial Navigation System

An Inertial Navigation System (INS) consists of accelerometers, gyroscopes, and magnetometers. For indoor navigation solutions these sensor types are mostly produced as micro-electro-mechanical systems (MEMS). The different sensor types of an INS are able to measure in all three space axis. With the recorded data of the accelerometer, the gyroscope, and the magnetometer it is possible to calculate the steric orientation and position. Short-term stability and the sensors bias require new filter models. To improve the accuracy of the measured orientation and position, Kalman-Filter can be helpful for estimating the position and orientation error (Haid *et al.*, 2004).

2.1. Hardware

For this project, an inertial measurement unit (IMU) was developed by the Competence Center for Applied Sensor Systems, which consists of two acceleration sensors, two gyroscopes, a magnetometer, a barometer, and a powerful ARM Cortex M3 processor as shown in figure 1. This board has been expanded with a Bluetooth interface that transmits measured data wirelessly to a PC. The board is operated by a lithium-polymer battery to operate independently. The processor performs the pre-processing of raw data coming from the sensors. The raw data is represented as digital units ("digits") and needs to be converted to SI-units for better visualization. Furthermore, the processor calculates the spatial orientation. For this purpose, the data of the accelerometers, the gyroscopes, and the magnetometer are merged using algorithms based on trigonometry. As a result, the spatial orientation in degrees for each sensor axis is transferred via Bluetooth.



Figure 1: Inertial Measurement Unit “CCASS IMU200” (Münter, 2011)

2.2. Software

The analysis and visualization of measured data is realized in MATLAB®. For this purpose, a graphical user interface (GUI), which is able to calculate spatial trajectories offline as well as online in real time, was developed. While moving the IMU, the IMU is visualized as a cube in a three-dimensional plot as shown in Figure 2. The cube changes periodically its actual position and its spatial orientation. For an off-line measurement, the data is stored in ASCII format. Using the ASCII formatted data the trajectory is calculated. This will facilitate the sharing of data between users. For online measurement, a virtual serial port is used. The measured data of the acceleration sensor and the angle of roll, pitch and yaw are comma-separatedly transferred.

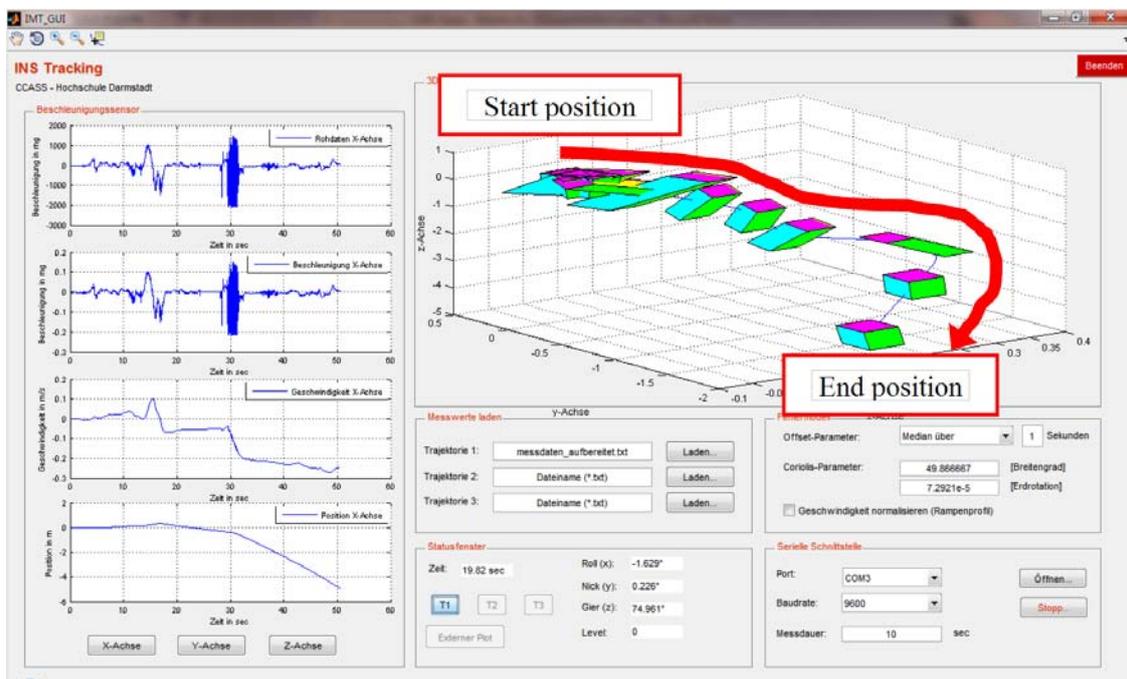


Figure 2: Graphical User Interface showing the trajectory of the IMU (Münter, 2011)

2.3. Algorithm

The MATLAB GUI receives the transmitted, comma-separated acceleration data and calculates the spatial position using double-integration and advanced filter algorithms. However, the accelerometer data contains errors. Thus, the gravitational acceleration is also recorded as static acceleration from the sensors, which can increase the square error in the position after a short time. To counteract, filters are necessary. Gravity has a direct impact on the acceleration sensors when the sensor axis is in the same direction vector of gravity. For this reason, the static offset, which is caused by the gravitational force, are compensated. The influence of gravity on each sensor axis is calculated by means of spatial orientation and acceleration data. That influence is subtracted from the acceleration data. The result is the

dynamic acceleration. Despite all, the dynamic acceleration still contains errors due to nonlinearity, linearity errors, and external influences such as temperature drift and accelerations caused by Coriolis forces. To compensate such influences and errors, a complex error model is required that takes part of all these factors and errors. Suitable filters are shaping filters and Kalman filter algorithms (Haid *et al.*, 2004). Both shaping and Kalman filters calculate the optimal error based on an error model and derive the probable location. The use of such filters can also improve short-term stability of the acceleration sensors. Despite the computational complexity, real-time measurement is possible. The visualization, however, requires more computational effort, so the visualization is separated in the test trials for the calculation of the position (Münter, 2011).

3. Application scenarios

In the following, three innovative and creative application scenarios for a three-dimensional communication interface between human and machine are presented. In general, the idea is to create hands-on models with sensor technologies instead of creating virtual models. Hands-on models might be useful in customer relationship, particularly in dialogues with customers.

3.1. Interior Design

A possible application scenario is found in the interior design. When setting up a house or apartment CAD programs are usually used for interior furnishings. Using a computer, the designer creates a model fitting to the customers' requirements. The customer can influence the model only by direct communication to the designer. Instead, the interior of an apartment can also be designed interactively with the customer. For this purpose, it is assumed that the designer has objects such as chairs, plants, and desks as miniature models. These miniature models are equipped with an inertial sensor system as shown in figure 1. The miniature models are freely movable and can be stacked on a multi-touch screen for example. Each model has a virtual model that is stored on a computer. When raising a model with built-in IMU from one floor to the next level, a virtual model can be visualized in real-time. Craftsmen can use this virtual model later as a construction print. The big advantage for the customer in this scenario is found in the interaction with the miniature model, because the customer can be creatively active. This means a reduced workload for the designer, since the three-dimensional communication interface is ideal for rapid prototyping.

3.2. Entertainment Industry

In the entertainment industry innovative games are possible that can take place two-dimensional as well as three-dimensional. Otherwise, it would be possible to develop a three-dimensional Memory[®] game or intelligent domino cubes that recognize whether they are toppled over or not.

3.3. Emergency Management

Major incidents provide always a great challenge for authorities and organizations with security tasks such as the fire and rescue services. Unclear conditions and a large contingent of rescue teams require an extensive emergency management. A multi-touch screen would help the incident command to screen relevant information about dispatched vehicles and the affected object. Emergency vehicles and objects could be realized as miniature models, which are equipped with an inertial sensor system. By placing a model of a vehicle on the multi-touch screen, the vehicle is assigned to the emergency and could obtain current information from the incident command.

4. References

Haid, M.; Marquardt, G.; Melander, S.; Nguyen, P. (2004): Verbesserung der referenzlosen inertialen Objektverfolgung zur Indoor-Navigation durch Anwendung der Kalman-Filterung, Sonderdruck aus VDI-Berichte Nr. 1829, pp805-808

Münter, M. (2011): Konzeptstudie zur dreidimensionalen Objektverfolgung und Identifizierung, Hochschule Darmstadt, Competence Center for Applied Sensor Systems (CCASS), Darmstadt.

Chapter 2

Media

Demoscene Computer Artists and Community

C. Hastik, A. Steinmetz

Department of Media
Hochschule Darmstadt – University of Applied Sciences, Darmstadt, Germany
e-mail : canan.hastik@h-da.de

Abstract

The overwhelming variety of subjects in the field of born-digital content makes it difficult to classify and establish digital creative artwork. The lack of historical distance makes it even more difficult to identify art movements and summarize new art forms. Analyzing the Demoscene, a European subculture having the roots in the field of early computer generated graphic art, provides a first approach to structuring the scene and establish demoscene art as a facet of digital art. This formal structure is the basis for a conceptual solution in the field of digital preservation of complex dynamic media objects.

Keywords

Computer Demoscene, Digital Art, Ontology

1. Introduction

This contribution is part of a research to ensure the improvement of the long-term preservation of complex digital artefacts and the knowledge transfer of digital handcraft techniques. It is about analyzing and structuring the technology used materials and methods to constitute the Demoscene as an art movement.

The field of real-time audiovisual animation as one facet seems to be the royal discipline in which creative individual performances of various artistic handcraft practices are combined. This makes these computer-generated presentations, their platforms and not least the scene primary objects of research. A sustaining documentation of this vivid and largely undocumented world of sub cultural “Demo art” requires a fundamental understanding of the origins of tools being used which further allows the examination and analysis of artistic and experimental use of media technology.

Domain specific research, data collection and analysis take a key role in the context of documentation. Based on archives, portals and community websites relevant topics, objects and their relations have to be analyzed to visualize the context. An explicit analysis and characterization of the creative handcraft by the example of “Computer Demoscene” is mandatory.

2. The Computer Demoscene

Together with technical development niche cultures arise, forming their own norms, values and specific practices like net art, pixel art or Demo art. To distill the defining aspects of the Computer Demoscene, a historical overview is helpful.

The roots of the Demoscene reach far back to the first computer generated graphic art subcultures in the early 60s. Ben Laposky in the USA and Herbert W. Franke in Europe are considered as pioneers in the field of early computer art (Goodman 1987). Laposky’s creations of fleeting light images using a cathode-ray oscilloscope by supplying the deflecting electrodes with varying voltages based on different time functions is similar to classical elements used in Demo art. The same goes for Franke’s experiments programming geometric elements and curves on analogue and digital technology. This also applies to Charles Csuri’s first real-time animations and the usage of computer technology as a medium for art (Csuri 2012). His programmed functions with attributes manipulated by mathematical instructions are as sophisticated as algorithms used in Demo art.

The Computer Demoscene began in the early 80s where programming became a popular hobby activity. Until the 90s the scene was closely associated with the cracker scene. Demo artists initially developed small introductory presentations for cracked home computer games. This digital signature, so called Cracktro or Crack-Intro, was a start screen with logo of the cracker group, colored text, marquee with information on the game and greetings to friendly cracker groups, graphics, music and effects using the background color.



Figure 1: Crack-Intro (Fairlight, 1987)

Soon these cracktros became more spectacular than the games and developed into independent, real-time graphics, motion graphics and audiovisual arts.

2.1. Demoscene Insights

To complement their skills Demoscene artists formed groups of programmers, graphic designers and musicians, so called demo groups. The graphic artist wants to show how good he is at creating pictures and textures, the musician want to show how great he can compose and the programmer, also called coder, wants to demonstrate how well he can fit all together and what technical programming tricks and effects he can get out of the given hardware. The goal is to put the audience in awe, to impress and entertain.

The Computer Demoscene with their Demo art is a creative subculture with its own artistic expression and scene specific language. It is defined as “aspiring computer artists everywhere” (Shatz, P. 1993), “all people interested in demos” (Kuittinen 2001), “the scene, the demo community – a worldwide community of hobbyists interested in computer demos” (Reunanen 2010) and “a worldwide network of computer enthusiasts... a sub culture of the home computer culture” (Bolz 2011) producing “real-time, non-interactive applications along with music and graphics” (Scholz 2007).

These applications are executable programs that typically represent real-time audio-visual animations. Several special forms of Demo art products originated and were basically first classified by Borzyskowski in 2000 as follows:

Intro:	One or two routines
Dentro:	Preview of a demo
Demo:	More than two routines
Mega-Demo:	Linking of several demos.

A sampling survey of pouet.net, the largest web repository of news, groups and productions shows that Demoscene artworks today are primarily categorized into “Cracktros”, “Intros”, “Demos” and “Wild” (Pouet 2000). The category “Mega-demo” does not exist and just a few “Dentros” are collected. In total more than forty thousand artworks could be counted by crawling the portal and following the links to the original resources and collecting them.

Demo art type	Number of objects
Demo	ca. 24 000
Intro	ca. 5800
Cracktro	ca. 4400
4k	ca. 2000
64k	ca. 2000
Wild	ca. 1800
Invitation	ca. 1000
256b	ca. 700
1K	ca. 400
Dentro	ca. 350
Procedural graphics	ca. 150
Other	ca. 1700

Table 1: Different Demo art types and number of objects

It seems that the classification defined by Borzyskowski is not established or has changed due to technological change. One possible approach to the description of the basic structures of scene objects in form of an ontology is outlined in the following graphic. Objects, properties and their relations are exemplarily represented to visualize the context of Demoscene tools and materials today.

Over the years productions were subdivided into several competition oriented categories based on size limits and platforms they are designed for. Competitions are held on Demoscene events which are the most important community meetings for presenting new releases. For each event general competition rules and categories were defined. These rules and categories are indicative for an important quality criterion and are regarded as a constructive challenge within the scene. The general rules defining these restrictions seem to be not standardized and are changing gradually. On top of that some products cannot explicitly be assigned to only one category, they are categorized by size or not at all.

By analyzing the defined categories and rules of recent and older events published on the event websites, three main categories can be identified: Demo, Intro and others like graphics, music, wild, game and sub categories. While Demos are mostly categorized by hardware platforms like Commodore C64, Amiga, PC and Atari ST, Intros usually are classified by size limits like 4 kilobytes and 64 kilobyte and sometimes also by hardware platforms.

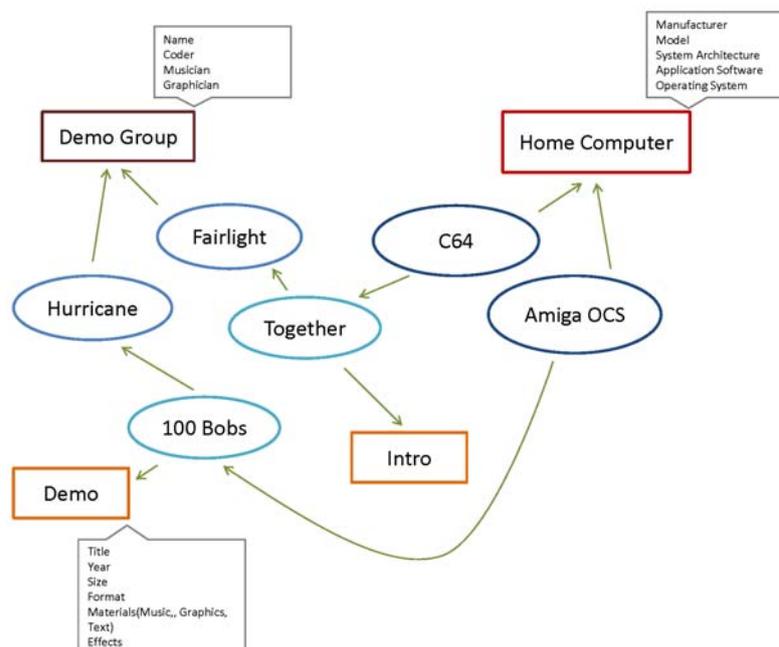


Figure 2: A model of Demoscene ontology

In total the use of over seventy different platforms can be counted, reaching from classic platforms like Commodore 64, Amiga and Atari ST to game consoles, handheld devices, mobile phones, operating systems and graphical user interfaces. Usually the activity is proportional to the actual distribution range of the platform, but also the access to appropriate development software plays a role.

2.2. Role of Hardware

First of all the “hackability” of the platform and its specifications formed the character of Demo artwork. Classic hardware with 8-bit, 16-bit and 32-bit architecture was commonly used.

Roughly speaking, with much effort not well documented or even inaccessible hardware was analyzed and reverse-engineered to create harmonically composed artwork with spectacular visual and audio tricks. Therefore a Demo programmer is often referred to as a craftsman who has mastered a particular cultural technique. The more simple the technological basis is, the more compact the net of aesthetic stimuli wants to be woven (Heikkilä, 2010) to realize always more complex and elaborate programmed tricks. It is expected that Demo artists will demonstrate their skills and pursue the principle to generate “... flashy bits written in custom assembly language and breaking every rules ...” (Shatz, 1993). In fact Assembler is still used for performance critical code but also programming languages like C++ are common. Beyond that there are various approaches of developing modular Demo editors. One example is mentioned at

this point, because it represents a reflection of the basic principles of the scene in dealing with resources and materials (farbrausch 2000).

From the impressive artistic artifacts that challenge the computer hardware at most, not only platform-specific styles but also compositional principles have emerged and are maintained until today (Hartmann, 2010).

3. Culture Techniques of Demo artists

The development of specific design techniques depends on the technical skills of the artist dealing with the machine, the tools and the scene specific handling of the existing repertoire of effects. Given hardware characteristics were successively studied and tested. However it can be observed that the use of new platforms will always build up on the use of an existing repertoire. On the one hand, active inventory, the backup and transfer of classical effects and principles of composition to new platforms is practiced. On the other hand, the new platform is used for more efficient implementation of established aesthetics and new versions of classic styles and principles arise.

3.1. Demoscene Classics

The structure of classic Demo art productions can be characterized by the use of classical elements which depends on the hardware. Graphics were animated with routines. Animations were made up of increasingly complex mathematically described objects and geometric shapes.

For platforms like C64 and Amiga OCS: text got scaled up and down, rotated, deformed, moved, fluttered and was typically presented in fast-paced or even dancing scroll effects or animation, tunnel, plasma, light and fire effects. Other classic old-school effects are for example the raster line interrupt and copper bar effect, both background effects that will display vertical and / or horizontal stripes of different resolution and color number on the screen. Programmers used clever tricks to make the elements look better and enriched them with as much effects as possible. Over time, highly complex effects and a variety of elements were used. While the classic Demo art also experienced a change in composition and content, Demos today are dominated by procedurally generated realistic 3D scenarios. With the widespread use of PCs in the mid 90s and the related variety of hardware a new era of the Computer Demoscene began. In contrast to the home computers, Demos on a PC may or may not work on another PC or are differently interpreting the program code. The Demo development was changing, screen composition, colors and innovative ideas came to the fore. Graphical presentation and fresh ideas had to convince the audience not only the technical masterpiece. Classic effects had to be reinvented or went out of fashion.

Faster processors and more computing resources were changing the possibilities previously limited by the hardware, for example playing a modest number of frames for a smooth, full-scale animation in the form of a film. The factor "real time" emerged as one most important quality criteria and principle of the Demoscene. Before real-time was the only way to animate images on the screen. For the viewer of a Demo it is not recognizable if a movie file or an executable program file with code-based animated graphics and sound is being played. He cannot monitor the real-time aspect. Real-time effects can only be judged based on the knowledge about the specific hardware requirements. If you record all Demos on video, it will not matter if one artist has released a Demo running ten seconds on ten floppy disks, or a Demo running 20 minutes on one floppy disk (Botz, 2011). By limiting the hardware and the size of the executable file not just comparability is achieved, but also the use of too many pre-computed animations is avoided (Reunanen 2010). These restrictions are regarded as a constructive challenge and are indicative for other important quality criteria of the scene.

3.2. Quantity and Tricks

Both the C64 and the Amiga only offer limited options to move objects horizontally and vertically on the screen. With this limitation a purely quantitative competition started.

Object to object records were broken, better written calculation routines became faster and more efficient. Same applies to image and sound productions. The available drawing programs for traditional platforms like for example Deluxe Paint for the Amiga generally provide all aspects of the graphics hardware. Users had access to a wide range of graphical features and effects limited to the original color graphics modes and the specifications of used hardware.

However, it was not uncommon improve them and make these add-ons available for free within the scene. Only through outstanding programming achievements and pioneering spirit existing limitations could be greatly exceeded and impressive graphics could be created. For example, special routines allow displaying up to 128 colors instead of the original 16 colors on a C64 screen (C64 Picture Gallery, 1999).



Figure 3: Hurricane – 100 Bobs, (Amiga-OCS-Demo, 1989)

4. Conclusion

The first analysis shows that the language used by the Demoscene community is formed by structural and social conditions but in fact the used language describes scene typical phenomena. The variety of subjects and forms of Demo art based on a huge amount of platforms using certain tools makes it difficult to develop a classification of this art movement and distinguish established trends in the Demoscene.

This research exemplarily outlines artistic works and practices as well as digital production techniques of computer generated visual media art from the early 80s until today. The analyzed portal contains a lot of valuable resources and context information as well as community annotations, but many resources are no longer available and especially the use of metadata is not uniform or even standardized. These inconsistencies in the specification of data structures complicate the accessibility to Demoscene media assets for public.

In further research more data and facts have to be collected and methodically processed while taking determined criteria for preservation (Hastik 2012) into account. A survey of all available internet resources must be made to develop a standardized metadata model.

5. References

- Borzyskowski, G. (2000), „The Hacker Demo Scene and it’s Cultural Artifacts“, <http://www.scheib.net/play/demos/what/borzyskowski/>, (Accessed 15 March 2012).
- Botz, D. (2011), *Kunst, Code und Maschine. Die Ästhetik der Computer-Demoszene*, Transcript Verlag, Bielefeld, ISBN:978-3-8376-1749-8.
- C64 Picture Gallery (1999), „A Brief Description Of Graphic Modes“, <http://www.studiostyle.sk/dmagic/gallery/gfxmodes.htm>, (Accessed 21 January 2012).
- Farbrausch (2000), „Fr-08: .the .product“, <http://www.theproduct.de>, (Accessed 15 January 2012).
- Franke, H. W. (1957), *Kunst und Konstruktion. Physik und Mathematik als fotografisches Experiment*, Bruckmann, München.
- Goodman, D. (1987), *Digital Visions: Computer and Art*. Abrams, New York. ISBN: 978-0810923614.
- Hartmann, D. (2010), „Computer Demos and the Demoscene: Artistic Subcultural Innovation in Real-Time“, in Funke, J. et al. (Ed.): *Proceedings of the 16th International Symposium of Electronic Art*, Revolver Publishing, Berlin, ISBN: 978-3-86895-103-5.
- Hastik, C. (2012), „Computer Technology- A Tool in the hand of the artist?“, in *Proceedings of Euromedia 2012*, Bukarest, Romania. (Accepted 02 March 2012)
- Heikkilä, V.-M. (2010), „Defining Computationally Minimal Art (Or taking the „8“ out of „8-bit“), <http://www.pelulamu.net/countercomplex/computationally-minimal-art>, (Accessed 15 January 2012)
- Institut für Mathematik und Informatik (2012), „Der Computer als Werkzeug der praktischen Kunst und der Kunstwissenschaft“, <http://stubber.math-inf.uni-greifswald.de/mathematik+kunst/computer.html>, (Accessed 15 January 2012)
- Klütsch, C. (2007), *Computergrafik: ästhetische Experimente zwischen zwei Kulturen. Die Anfänge der Computerkunst in den 1960er Jahren*, Springer, London, ISBN: 978-3-211-39409-0.
- Kuittinen, P. (2001), „Computer Demos – The story so far“, <http://mlab.uiah.fi/~eye/demos/#glossary>, (Accessed 15 March 2012)
- Pouet (2000), „Your online demoscene ressource“, <http://www.pouet.net/> (Accessed 15 March 2012)

Reunanen, M. (2010), *Computer Demos – What Makes Them Tick?*, Licentiate Thesis, Aalto Univ., Helsinki.

Serexhe, B. (2011), *Substanz und Ethik in der Konservierung digitaler Medienkunst*, ICOM Deutschland. Mitteilungen 2011, Vol. 18, No.33, pp8-10.

Serexhe, B. (2012), *Digitale Herausforderungen*, *Digital Art Works: The Challenges of Conservation*, pp4-8.

Scholz, A. (2007), *Iconoclash: Opium for the masses*, in SCEEN magazine, no. 2, pp. 51-56.

Shatz, P. (1993) *Walkthroughs and Flybys*, Waite Group Press, Corte Madera, ISBN: 1-878739-40-9.

Cognitive prototypes and narrative thinking

P. Green

School of Art and Media
Plymouth University, Plymouth, United Kingdom
e-mail: paul.green@plymouth.ac.uk

Abstract

The emergence of interest in 'experience' over 'use' in interaction design has recast the role of the user from a 'cog in rational machine' to one who experiences technology as part of a living environment (McCarthy and Wright, 2004). This shift in emphasis is part of a longer discourse in HCI which charts a trajectory from expert user to social actor (Grudin, 1990; Bannon, 1991). As part of this shift from *usability* towards *experience* there has been an increasingly visible presence of an artistic attitude in a field of HCI. While the broader context of this research is concerned with the design of experience around responsive artefacts this paper concentrates on how 'narrative thinking' operates within the context of open-ended visual content. It promotes the decoupling of narrative from material artefacts and emphasises reader centric perspectives which hinge on personal experience and meaning making. The paper represents one step in an argument for establishing a narrative framework for creative practice with a particular interest in responsive artefacts.

Keywords

Narrative, visual content, cognitive narratology, theory

1. Introduction

While narrative has been the focus of some considerable attention in the arts, emphasis has largely been placed on media seen as best affording the sequential unfolding of a story. Those that have attracted commentary in relation to narrative have typically been media that implicitly support sequences of visual imagery such as comic strips, film, and animation. The current overlap between narrative and interaction is probably most evident in the field of game studies and the Interactive Digital Storytelling (IDS) community which fused the interests of Technologies for Interactive Digital Storytelling (TIDSE) and International Conference on Virtual Storytelling (ICVS) conference series. With the emergence of digital media practices there has been more explicit openness towards a concept of narrative which is user/reader centric and many strands of narrative discourse have emerged around topics such as intelligent agents, tools for authoring, user/player agency and the blending of concepts of agency and authorship. There has also been a significant increase in applications focussed on the benefits of interactive narrative for learning demonstrated through, for example, the work of the Kaleidoscope network. By and large these approaches focus on interactive artefacts as containers of stories with some exceptions which include mobile or situated storytelling evidenced through projects such as 'Murmur' initiated in 2003 or 'Remember Me' (Speed et al., 2010). The last example explores approaches implicating the artefact as a conduit for memories where objects tagged with QR codes offer concrete material evidence of events and experiences recounted by their owners. Here the artefact is a witness to events rather than a medium that actively narrates as a film or novel might do. Narrative can therefore be treated as offset from the material object.

While the wider scope of the research involves unpacking narrative experience in a social spatial context of responsive or interactive objects. This paper is a step towards that and focuses on decoupling narrative from artefacts. To this end it explores how a cognitive approach is one significant element that helps support a broader concept of narrative as a framework for creative practice.

2. Context

There are two observations with regard to recent approaches to narrative that help to provide some focus here about how it is possible to understand how 'narrative thinking' operates with respect to visual content. These observations come from outside of the discipline of visual art and relate to research in literary fiction and the social sciences.

First, within contemporary discourse involving narrative across media the selection of material for analysis seems often based on its suitability to extend existing concepts and metaphors into the discussion of new media technologies. Aaron Aarseth in particular has illustrated the way ill defined labyrinthine metaphors have been uncritically projected onto new technologies (1997, p. 7). In art and aesthetics discourse the emphasis on structural, or 'grammatical', features relating to *the plot*, derived in part from Saussurean linguistics, has affected a heavy constraint on identifying suitable media for analysis. This is despite ideas that emerged almost immediately on the heels of Francophone structuralism in

the 1960s and emphasised reader/audience centred constructivist perspectives on narrative. In literature for example interlinguistic and intertextual concepts in the work of M.H. Bakhtin and Julia Kristeva, as well as reader centred ideas from Roman Ingarden and Wolfgang Iser, in particular gave credence to the constructive role the reader played in actualising stories that were 'virtually' available in a text. In addition, those working more recently in the field of *cognitive narratology* recognise how certain strands of research within cognitive psychology can offer support to how we understand narrative in literary form. Such moves towards marshalling cognitive theories in support of understanding techniques employed in literary works also offer themselves favourably to an understanding of narrative in interactive and visual media.

A second observation is related to how the *narrative turn* in the social sciences has accumulated momentum in the last two decades. The attention to personal *small* stories relative to *canonical narratives* in this field implies a different concern with what narrative can offer as a method for generating knowledge or understanding. Within areas such as discourse and identity studies, for example, there is a particular concern with how narrative can support enquiries about cultural difference or the construction of personal identities. More generally contemporary narrative research in the social sciences often focuses on how we go about attributing meaning and significance to events in our daily lives. While arguably such research has its roots in a sociolinguistic strand of narrative research which emerged synchronously with the classic structuralist narratology, it has wielded far less influence over practitioners working in the visual arts. The attribution of meaning and significance to aspects of lived experience, which is a core feature of narrative theory in social studies, would appear to hold much benefit for contemporary media artists.

The following sections of the paper outline two important approaches in cognitive narratology before going on to discuss a particular case for visual media. The narrative turn in social practice mentioned above is not separately addressed but is instead inflected in the cognitive approaches outlined.

3. A case for cognitive narratology

3.1. Binary and scalar approaches to narrative

To help elucidate the case for applying a cognitive approach to narrative in visual media it is useful to illustrate a common methodological division where one view focuses on a *binary* or *situated* definitions of narrative while another takes narrative within a *scalar* model which allows for the transferability of stories across a variety of media. The first might hold to a position that versions of stories when moved through history or across media cannot be considered the same stories as their value and meaning is located in the situation and context of their telling (Smith, 1980, Aaron, 2009). One might also face contentions about the narrative status of photographs or paintings based on a premise that a minimum of two events must be cited as a necessary condition for narrative (Carroll, 2001, p. 119). The latter view looks for the ways in which essential narrative properties are variably manifested in different media and is the basic "hypothesis of classic structuralist narratology" (Herman, 2002, 2004 p. 51). Here a novel might be regarded as possessing more narrativity than a painting since it can demonstrate more of the essential properties of narrative (Ryan, 2006, p7-9). One might also find arguments for how certain media can enable narratable content - for example, the inner thoughts of a character cannot be explicitly represented in a painting but can in cinematic production with an omniscient narrator.

From the perspective of visual media the problem with the first view is that it does not account for what might generally be thought of as narrative features in the image in Figure 1a. If narrative requires the depiction of two events this image does not appear to meet the basic criteria and fails to qualify as a narrative. Nevertheless it is hard to deny there is some quality of narrative that appears to be present in the image. In order to argue the case for narrative here it is then necessary to begin identifying what qualifies as an event and whether events need to be directly depicted in the frame of the image. While such analytic debates do occur in the literature, particularly around art and aesthetics – see Bence Nanay (2009) - the focus of the arguments often become political ones that defend the status of visual media. To shift focus away from the medium and instead emphasise the relationship or effect of the text on the viewer, reader, or audience, and the role the recipient(s) takes up in relation to participating in the construction of the content underlines contemporary approaches that are captured within the field of cognitive narratology.



Figure 1

3.2. Preference rules and ‘natural’ narrative

Researchers working in this tradition come from various disciplinary backgrounds including cognitive psychology, social psychology, cognitive linguistics, discourse & conversation analysis, and contemporary literary theory. Of particular importance is the work of David Herman and Monika Fludernik. Herman has addressed concepts from cognitive science within the context of modern and contemporary cultural studies and literary art (2002, 2003, 2004, 2010), and Fludernik’s efforts towards a ‘natural narratology’ (1996) are especially relevant in laying some of the ground for how cognitive approaches can be usefully applied in the context of visual and interactive media.

Herman is a leading figure in drawing together the argument for the application of cognitive science approaches to literary art. He shows how the field can be divided into major strategies that on the one hand focus on *making sense of stories* and on another underline a *sense making strategy*. While considerable crossover occurs between the interests of those researching in each of these strands one can summarise by saying the first deals with the way recipients of stories mentally articulate what is going on by interpreting specific cues in a text and the latter shows how stories can be used as tools for making sense of the world. The second strategy takes form in areas such as cognitive therapy, psychoanalysis, social and cultural studies of race and gender. The argument here for narrative in new media relates to how visual and interactive media can be processed in narrative terms and does not constitute an effort to create a new expanded definition of narrative which can suitably cater for visual and interactive media. Rather it makes a move towards exploring how narrative might be utilised as a method in creative practice by drawing on contemporary cognitive approaches that contribute to, what Herman calls, a ‘second cognitive revolution’ (2010, p.156). He defines a first cognitive revolution as a reaction to Behaviorism where the mind was viewed as ‘software’ running on the ‘hardware’ of the brain. This first revolution was constituted mainly by research in cognitive psychology and early artificial intelligence. The second revolution is concerned with a view of the mind which can be understood as intertextual, situated, relative, distributed, intermental, and social and takes in a much broader scope of disciplines that span science, art, and the humanities.

While Herman champions the second cognitive revolution he does not lose sight of the importance of classic structuralism. In *Story Logic* (2002) he outlines a preference based typology for action in narrative, reproduced in Figure 2, and discusses this in relation to different genres. This exemplifies the scalar approach, mentioned above in section 3.1, and helps to illustrate how established genres provide expectations about what constitutes action and how action is likely to be distributed in narratives of a given genre.

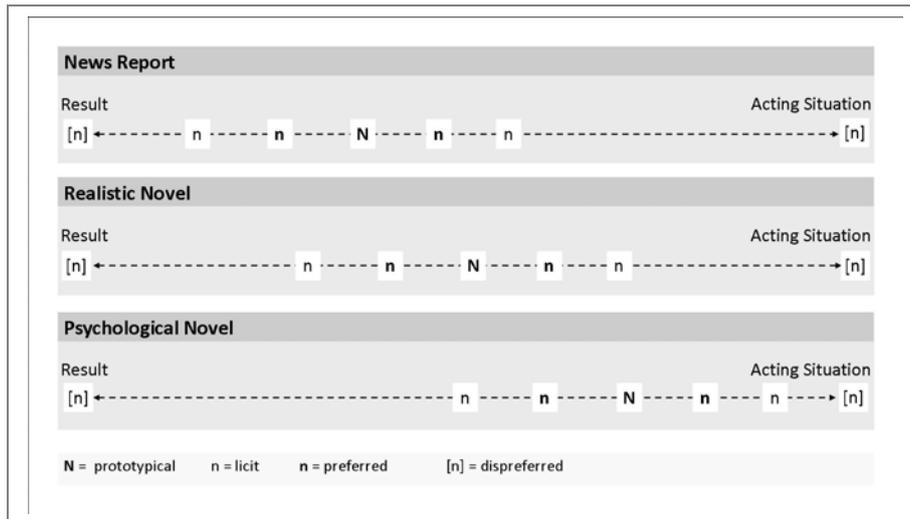


Figure 2: Typology of preference rankings for action representations in narrative as presented by Herman (2002, p.60)

The diagram represents three narrative genres with the preferred distribution of action prototypically placed between two poles – one pole focussed on *result* and the other on the *acting situation*. In this case ‘result’ indicates a tendency towards wrapping action in clear causal logic. The *acting situation* on the other hand is a term attributed to Georg Henrik Von Wright where a complete description of actions requires: ‘a’ an initial state before the action is initiated; ‘b’ an end state after the action is complete; and ‘c’ the state of the world had the action not occurred. This acting situation (which is the *possibility* for action between ‘a’ and ‘b’) weighs on the ability to compare the world as it is to some other counterfactual state that might have existed otherwise. The acting situation therefore can be understood as all the possibilities in between ‘a’ and ‘b’ that could rationally explain the event(s) that are explicitly available in a text – or image. The tendency towards open endedness in certain genres, such as the murder mystery or psychological novel, supports the aesthetic affordances of the genre which involves allowing the reader in to participate in constructing what it is that’s going on. Frank Kermode’s ‘A Sense of an Ending’ captures this participative activity of the audience in always looking forward from the current situation described in a story and attempting to narrow the possibilities for how it might end. In essence Herman shows there is no universal formula per se for how action should be depicted in all narratives whether we survey it synchronically across genres, as is the cases in Figure 2, or take a diachronic approach over generations of literature, as Fludernik did in her influential “Towards a ‘Natural’ Narratology” (1996). This text has served as a significant support to the emergence of the current thinking around cognitive narratology in relation to literary theory.

Like Herman, Fludernik does not see the cognitive turn in narratology as supplanting the ideas that arrived in the 1960s through Francophone Structuralism. Instead she sees a cognitive framework as being sufficiently expansive to account for postmodern narrative works that sat on the experimental or end of the spectrum. (2003, p. 264). The cognitive groundwork for her concept of a ‘natural’ narratology, which is useful to consider within the context of visual and interactive media, draws on three different sources of the term ‘natural’. First is from a sociolinguistic tradition inspired by the groundbreaking work of William Labov and Joshua Waletzky (Labov, 1972) where natural narrative is associated with oral accounts of life experiences. Labov and Waletzky’s method paired with Garfinkel’s ethnomethodology (1967) significantly impacted on the development of conversation and discourse analysis and has been one of the major influences of contemporary interest in study of personal narrative. A second influence on Fludernik is in the area of cognitive linguistics that overlaps with prototype theory. Here a cognitive model of a bird is more likely to be represented as a crow or a sparrow as opposed to ostrich. Prototypical objects hence appear more ‘natural’ compared to other objects in the same category that are ranked less typical. Here Fludernik translates this cognitive notion of ‘naturalness’ to narrative by defining narratives of everyday experience - understood through exposure to contemporary fiction, news reports, or oral stories in conversation – as the prototype rather than those, for example, from previous generations or other centuries. Thirdly Fludernik was influenced by Jonathan Cullers notion of naturalness located in how readers blend or synthesise inconsistent information presented to them in a text. For example, rather than rejecting contradictory character behaviour or narration as implausible, a reading or viewing audience more often show a willingness to normalise the character’s behaviour by filling in gaps in logic or action. Fludernik uses Ishiguro’s butler Stevens in ‘Remains of the Day’ to illustrate this concept and points to the ‘unreliable narrator’ as an instance of this type of reading strategy (Fludernik, 2003, p. 251). In both of the approaches underlined by Herman and Fludernik above factors outside of the text contribute to how information explicitly presented in the work gets remoulded by an active reader.

3.3. Narrative gapping and significance

Narrative as a mode of human communication has been recognised as demarcating and reflecting *significance* in human experience. A key constituent of William Labov's theory, for example, is *evaluation* which is identifiable as a linguistic structure in talk and functions to orient the listener in a conversation towards the *point* of the story - why it is worth telling, or why it was a significant experience from the perspective of the narrator (Labov, 1972, p.366-75). As far as prototype theory goes the significant image (e.g. crow) of a given category (e.g. birds) referenced in a story is the particular cognitive image that surfaces in a readers/viewers/listeners consciousness. In narrative we can also recognise a nexus of prototypes in any given scene. For example, in a stereotypical bar fight which descends into chaos and involves characters breaking furniture over each other, we are unlikely to have an initial image of the item of furniture being IKEA bookshelves. Based on prototypes of bar fights we might be familiar with, such as in Westerns, an inventory of weapons will include bottles and glasses half full, and furniture will usually be constrained to chairs and possibly small bar tables. Any shock or humour we might experience from the representation of such a scene will often involve veering away from, while at the same time referencing, the prototypical inventory of weapons. Significant elements that support the construction of narrative therefore may be referenced but not necessarily made explicit in a text or image. Seymour Chatman's account for instance notes how *selection* is one of two principal features of narrative and is described as: "the capacity of any discourse to choose which events and objects to actually state and which only to imply" (Chatman, 1978, p. 28). While gapping is a necessary feature of communication in the narrative arts the gaps in action follow preference rules for different genres, as suggested by Herman, and are also naturalised, in Fludernik's sense.

Regarding independent still images, an untitled image in a gallery for example, the gaps are in fact what are salient. What it is that is not stated becomes the significant feature, the piece that the reader or viewer must themselves supply. Among a range of techniques noted in the study of narrative painting, for instance, the *punctum temporis* isolates a slice of action or moment in history which allows the viewer to unfold the possible events backwards and forwards from the moment in time depicted. Examples of neoclassical painting, such as that by Jacques Louis David, are often selected by aesthetic theorists to demonstrate the concept at work in still images (Nanay, 2009, Steiner, 2004). And while one could argue that such a salient visual moment does not 'tell' a specific story as a text might do, a cognitive narratological perspective can account for it by holding up what is 'natural' to a genre and how it offers up particular prototypes from which the viewer can extrapolate.

4. Case study

This case study attempts to ground some of the above concepts in a short discussion of two images which were exhibited together as a diptych (Figure 1b). In their original context the images were framed and hung side-by-side in a gallery. The two images might be said to be connected on a number of different levels. First of all they exist synchronously in front of a viewer and therefore are physically associated in spatial terms. One can also say they are stylistically or aesthetically associated in the sense they conform to one style of photography. They appear to be produced without any obvious postproduction and the lighting is similar in both. The images contain certain cues - related to the character clothing and posture as well as details in the environment - which would suggest the person depicted in both images is the same individual. If we can assume it is the same person in both images it is possible to proceed to further inferences about the time difference between the images and the role of the photographer. The handcuffs suggest not only some illegal action but also the prisoner is possibly in transit and therefore unlikely to be unaccompanied in this situation. Before moving to construct probable 'cause' about what she may have done, it is clear that we have already made a number of assumptions since there is clearly no firm evidence that these images involve the same individual. It does seem reasonable, or at least possible, to continue on a deeper process of reasoning drawing on the emotional expression of the subject to infer a particularly troublesome attitude. Our judgement of the subject hinges on whether the expression is read as *confrontational in the face of authority* or *resilient in the face of injustice*. Either way the reading is bound to produce different paths of consideration that articulate the most prototypical explanation of what it is that is going on.

To avoid make the above assumptions about these images they were presented over three sessions to small groups of subjects totalling 28 participants. Without any prior notice they were given approximately 2 minutes to write about what it was they believed was going on in the images. No other information was provided such as a title or information about the original context of the images in the gallery setting; neither was any clue provided as to whether they were to be integrated into a publication, or for what reason the images were produced. The images were projected in a lab in full screen presentation mode shielded from the computer desktop environment. In each session the written statements were collected immediately afterwards and some questions were put to the group as to why a particular statement such as 'she is being arrested' could be justified. This discussion went on for approximately 5 minutes after the statements were collected. The majority of subjects indicated that 'she' was 'being arrested'. When asked how they could know this all noted the handcuffs. The shirt was explicitly singled out as evidence that the images represented the same individual

despite the large degree of colour distortion between the images caused by tungsten lighting at the time of the shoot. When asked how they could know that the handcuffed character was a female, one subject insisted it was the curvature of the torso while others indicated the connections to the character in the left panel made due to the shirt. When asked why she was being arrested one subject volunteered that it was something ‘bad..a stabbing?’ and this was partially qualified by the ‘dirty look.. she gives you’. The reason given for such a specific association was that “..she looked like someone previously seen on the news..” who had conducted such an act. Another response to the question about the crime was that she was abroad and that it might be “..something to do with drugs.” The rationale offered for the girl being abroad was that she was tanned, the blinds in the background looked like they were from a beach hut and the plants seemed exotic. In addition the association with a particular recent TV documentary series about people being incarcerated in foreign countries was directly referenced as influencing this interpretation.

The rationale provided by the participants illustrates the extent to which the concepts of ‘prototyping’ and ‘natural narrative’ account for the construction of fictive events not depicted in images. In practice it is possible for subjects to extrapolate out from an open ended sign such as a facial expression to a very specific proposition about an action that is in fact not depicted. While such a proposition might not be successfully defended in factual terms this has no bearing on what a viewing subject may experience or imagine to be happening since narrative logic does not depend on empirical causality. The limited information explicitly depicted in Figure 1b allows us to draw on life experience and knowledge about our own world to ‘naturally’ infer what it is that is going on. Marie-Laure Ryan, who has made significant moves in addressing narrative across media, including visual and interactive works, might explain this through the ‘principle of minimal departure’. This refers to how, when we are exposed to a world represented in an artwork we invariably imagine it to be a mirror of our ‘actual’ world until some detail of the text or image contradicts our assumptions. Ryan holds that we insert our own descriptions of prototype worlds, entities, actions, events, and social behaviours until the evidence in the work breaks the prototype and forces us to engage with an alternative ‘possible world’. This creative intervention in, or subversion of, the prototype is what makes a work engaging; without such interventions we are left with cliché. And while cliché is recognisable and in a design context supports ‘usability’, it is not necessarily useful for producing significant ‘experiences’ for viewers.

5. Conclusion

In this paper I have presented a case which favours decoupling narrative content from the material artefacts. I have attempted to demonstrate how a cognitive approach to narrative, illustrated primarily through the work of David Herman and Monika Fludernik, can be support such a view. By displacing the artefact as a centralised container of narrative and instead exploring how it makes use of cues that trigger narrative thinking in subjects it is possible to extend the study and discussion of narrative to situations that are generally excluded from narrative enquiry. This perspective is one level in a broader context of investigation which explores how narrative logic can be useful as a support to the design and development of spaces that incorporate static and responsive artworks. In this way, through making use of narrative methods currently being applied in other disciplines, we may get closer to understanding how creative practitioners use the prototypical social, or ‘natural’, world as a resource for constructing engaging and memorable experiences.

6. Acknowledgements

I wish to acknowledge Crawford College of Art and Design and the CIT Research Office for the award of a grant which has supported the wider research associated with this project.

7. References

- Aaron, S. (2009), “Story Identity and Story Type”, *The Journal of Aesthetics and Art Criticism*, 67, 5-13.
- Aarseth, E. J. (1997), *Cybertext : perspectives on ergodic literature*, Baltimore, Md., Johns Hopkins University Press.
- Bannon, L. (1991), “From Human Factors to Human Actors: The role of Psychology and human-computer interaction in system design”, In J. Greenbaum and M. Kyng (eds.), *Design at Work: Cooperative Design of Computer Systems*, 25-44, Hillsdale, N.J.:Erlbaum.
- Carroll, N. (2001), *Beyond aesthetics : philosophical essays*, Cambridge, UK ; New York, Cambridge University Press.
- Chatman, S. B. (1978), *Story and discourse : narrative structure in fiction and film*, Ithaca, N.Y., Cornell University Press.
- Fludernik, M. (1996), *Towards a 'natural' narratology*, London ; New York, Routledge.
- Fludernik, M. (2003) “Natural narratology and cognitive parameters”, in: Herman, D. (ed.) *Narrative theory and the cognitive sciences*. Stanford, Calif.: CSLI Publications.

- Grudin, J. (1990), "The computer reaches out: the historical continuity of interface design", in *Proceedings of the Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people* (Seattle, Washington, United States, 1990). ACM.
- Herman, D. (2002), *Story logic : problems and possibilities of narrative*, Lincoln, Neb, University of Nebraska Press.
- Herman, D. (2003), *Narrative theory and the cognitive sciences*, Stanford, Calif., CSLI Publications.
- Herman, D. (2004), "Towards a Transmedial Narratology" in: Ryan, M.-L. (ed.) *Narrative across media : the languages of storytelling*. Lincoln: University of Nebraska Press.
- Herman, D. (2010), "Narrative theory after the second cognitive revolution" in: Zunshine, L. (ed.) *Introduction to cognitive cultural studies*. Baltimore: Johns Hopkins University Press.
- Labov, W. (1972), *Language in the inner city; studies in the Black English vernacular*, Philadelphia,, University of Pennsylvania Press.
- McCarthy, J. Wright, P. (2004), *Technology as Experience*, MIT Press.
- Nanay, B. (2009), "Narrative Pictures", *The Journal of Aesthetics and Art Criticism*, 67, 119-129.
- Ryan, M.-L. (2006), *Avatars of story*, Minneapolis, University of Minnesota Press.
- Smith, B. H. (1980), "Narrative Versions, Narrative Theories", *Critical Inquiry*, 7, 213-236.
- Speed, C. (2010), *Remember Me*. <http://www.futureeverything.org/festival2010/rememberme>.
- Steiner, W. (2004) "Narrative Pictures", in: Ryan, M.-L. (ed.) *Narrative across media : the languages of storytelling*. Lincoln: University of Nebraska Press.

Chapter 3

Business

Managing the Future Energy Policy for Ireland: Examining the Role of Nuclear Power

D. Lynch, A. Wright

Cork Institute of Technology, Cork, Ireland.

Abstract

This study assesses the attitudes and opinions of Irish people towards the use of nuclear power, and specifically examines the concept that Ireland should use nuclear power to generate electricity in the future. Currently, Ireland has a prohibition on nuclear power, with little debate or discussion on same, and with scant amounts of information published on the subject as it relates to Ireland. Conversely, from an international perspective, there is a significant volume of literature available covering all aspects of nuclear power, generated from over 50 years of nuclear power operations.

Nuclear power technology is perceived as being a relatively cheap source of electricity, with some positive environmental credentials. The evidence from this study, however, highlights that the public's attitudes are predominantly influenced by safety concerns and issues surrounding nuclear waste management. This research study has found that Irish people do not support nuclear power at present, but may consider supporting nuclear power under certain circumstances and conditions in the future. This current investigation reveals that if nuclear power could be proven to provide lower electricity costs, lower carbon dioxide emissions and enhanced security of supply, attitudes and opinions could change towards nuclear power to the point that it could become favoured.

This research has identified key findings that point to a desire that relevant information on nuclear power be made available in Ireland, and also for the instigation of a national debate on the use of this technology for electricity generation in the future.

Keywords

Nuclear Power, Climate Change, Future Energy, Government Policy and Prohibition on Nuclear Power.

1. Introduction

1.1. Energy Challenge

This research is an assessment of public attitudes and opinions towards the use of nuclear power in Ireland for the production of electricity and explores circumstances under which nuclear power may be utilised in the future.

The Energy Information Administration (2009) reports that world marketed energy consumption is projected to increase by 44 percent from 2006 to 2030. (Gittus, 2007) commented that forecasts reveal that in the 21st century, the world's reserves of coal, oil and gas will be virtually exhausted, and illustrates that Irish electricity generation by fuel type will underpin the reliance on fossil fuels. (Campbell, 2000) states that the discovery rate for crude oil has dropped to such a significant level that for every one barrel of conventional oil found, four barrels are consumed. (Gallis, 2009) states that most European countries are heavily reliant upon imported energy, with European Union (EU) countries as a whole importing 50% of their energy needs, a figure expected to rise to 70% by 2030. (Doran, 2009) further argues that if the EU is to secure a viable energy future, member states must develop a robust policy response based on the framework of collective energy security.

According to Sustainable Energy Ireland (2007), since the mid-1990s, Ireland's energy import dependency has grown considerably, from a figure of 69% in 1990, to a significant 90% in 2006. (King, 2009) argues that as Ireland relies on coal, oil, gas or peat for about 90% of its electricity, and only produces a tiny fraction of what is consumed. Ireland is now more heavily dependent on imported oil for energy requirements than almost every other European country.

1.2. Nuclear Power Option

According to (McWilliams, 2006), the world and Ireland is at a once-in-a-century crossroads in terms of energy, with a movement away from carbon based fuels to nuclear power. (Keenan, 2008) reported that Ireland, by building just one nuclear power station, could reduce carbon emissions from the electricity industry by a fifth. (McWilliams, 2006) states that Ireland simply has no alternative as oil is running out and as the regimes that control oil become increasingly unstable, oil supplies might be unsound even before it runs out. According to (Gittus, 2007), Ireland needs alternatives

energy sources if the country is to avoid shortages and 'Energy Wars', but these alternatives must be secure, environmentally acceptable and affordable.

(Quinn, 2009) further argues that the money spent on four years' worth of carbon emissions credits, (which is just under 1 billion Euro), would cover the cost for Ireland to build a medium-sized nuclear power station. According to (King, 2009), antipathy to nuclear power in Ireland has become a kind of secular religion, based almost entirely on superstition and bad science. (McWilliams, 2005) states that people in Ireland are afraid of the N (nuclear) word, so much so, that it cannot be used in polite conversation, being a word so vile, foul and degrading, it automatically puts you 'outside the pale'.

In this research, three general circumstances have been reviewed and discussed: first, the commerciality of nuclear power and its potential impact on electricity prices in Ireland, second, nuclear powers environmental credentials and its potential contribution to lowering CO₂ emissions, and third the enhanced security of supply that nuclear power may bring. These three criteria or set of circumstances are the main focus of this research, which explores whether one, or a combination of the three, could lead to Ireland removing the prohibition on nuclear power, and building its first nuclear powered electricity generating station.

2. Research Methodology

This research is based on the hypothesis that if the Irish public were provided with more information on the potential benefits of nuclear power, and such information was founded on governmental, international and institutional evidence supporting the commercial competitiveness, environmental friendliness, and enhanced security of supply of nuclear power, a significant increase in public support for the use of nuclear power in Ireland could be achieved. The theory underpinning this research study is that public attitudes and opinions in Ireland towards nuclear power are directly proportional to the amount of information available on the subject, and the amount of public debate on such information. This research is intended to achieve a number of objectives:

- To examine attitudes and opinions towards the current usage of energy in Ireland for electricity production.
- To understand attitudes and perceptions towards lowering Irelands CO₂ emissions.
- To investigate if the use of nuclear power in Ireland would be considered acceptable based on the following three criteria:
 - Commercially Competitive
 - Environmentally friendly
 - Security of supply enhancement
- To examine if there was a solid platform for a nuclear power program to be debated in Ireland.

The researcher chose to use a quantitative methodology, employing a survey, to collect data to generate statistical information in pursuit of finding answers to the research question. The decision to choose this method was heavily influenced by the researcher's belief that quantitative research would generate more reliable data and information over that which qualitative research would provide in answering the research question. (Cohen, 2000), states that researchers often favor questionnaires as all the participants are exposed to the same questions.

This research study was undertaken to investigate amongst a cross section of Irish people general attitudes and opinions towards nuclear power, and to determine if these attitudes and opinions would change if circumstances arose where commercial, environmental and enhanced security of supply benefits could be proven to accrue from the use of nuclear power for electricity generation in Ireland. The researcher set thirty two questions for participants to answer. Questions one to fourteen focused on attitudes and opinions as they relate to the fuel mix used in Ireland to generate electricity. Questions fifteen to thirty two focused more on nuclear power.

The aim of the survey was to generalise the voting population of Ireland. To achieve this aim, the researcher chose the simple random sampling method, targeting adults over the age of 18 to complete the survey. The researcher initially distributed the survey to a select group of known people, who were then asked to randomly distribute the survey to adults over the age of 18 years. The researcher chose this method as a means of avoiding bias in the sample. In order to minimise the risk of researcher induced bias, the researcher had no direct contact with any of the final selected survey

participants and therefore could not offer any instructions on the completion of the survey. In terms of survey participant selection, the only instruction given on control of randomly selected participants was that each participant must have been 18 years old or older before they could complete the survey. This single criterion for selection of participants was chosen as it is the lower legal age limit for voting in Ireland. This age limit and its corresponding legal right to vote was heavily influenced by fact that changes to policy via a referendum on the use of Nuclear Power in Ireland would only be open to persons of the voting age. No other limiting criteria would apply in a referendum vote situation.

The limitation of this research was the lack of good quality information relating to nuclear power in Ireland. A further limitation of this research was the difficulty in reaching a larger sample size. The researcher would have preferred to have worked with a larger sample size of one hundred or greater. The final sample size in this research, eighty participants, was limited by time and complexity of administrating the survey questionnaire.

3. Literature Review

3.1. Nuclear Power: An Irish Context

Ireland has no history of nuclear power electricity generation, with the use of nuclear power prohibited under legislation in Ireland. Therefore, there is no history to base a literature review of the commercial, environmental or security of supply aspects of nuclear power in Ireland. This literature review presents experiences from countries that have had a history of operating nuclear power plants. (Ahern, 2007) wrote that *never in our history has our energy policy been so important, not only due to its role in fuelling the engine of the economy, but also given its centrality in how we manage and protect our environment and respond to climate change. In this context, energy policy and environmental policy are seen as two sides of the same coin*, (2007: 3).

According to the Department of Communications Energy and Natural Resources (2007), there is a need to focus on Ireland's demand and use of fossil fuels for the production of energy, on the use of renewable energy sources, setting of targets for a reduction in the use of carbon dioxide (CO₂), generating fossil fuels, promotion and support for a greater dependence and usage of renewable energy technologies and the Kyoto based targets for limiting the emissions of carbon dioxide, amongst other things.

3.2. Prohibition on Nuclear Power in Ireland

According to (Ryan, 2007), the Irish government believes that nuclear power is neither sustainable nor the answer to Ireland's energy needs. According to (McManus, 2007), there is no percentage for Ireland to get involved in the debate on nuclear power, politically and legally it is a 'non-runner'. Irish people have enough to be doing without wasting their time on having a debate on nuclear power, (McManus, 2007). The Irish government fully intends to maintain the statutory prohibition on nuclear generation in Ireland and confirmed that the Department of Communications, Energy and natural Resources will not be undertaking a feasibility study on the construction of a nuclear reactor (Ryan, 2007). According to (Turvey *et al.*, 2008), whilst noting the prohibition on the use of nuclear power in Ireland, states that the whole issue of nuclear power in Ireland is a taboo, and believes that nuclear power is one option which should be debated openly in Ireland, with the facts on nuclear power being freely available.

According to the (Turvey *et al.*, 2008), nuclear power is essential to Ireland's future well-being, as it is highly-competitive in cost, taking into full account the cost of waste disposal and plant decommissioning. As noted by the International Energy Agency (2007), the cost of a unit of electricity generated from nuclear power is shown to be on a par with that from coal; cheaper than other fossil fuels and considerably cheaper than wind, which when one considers that as the prices of oil and fossil fuel increase, the economics of nuclear power get better and better. (Kelly, 2008) notes that on a global level, nuclear power plays a role in the large-scale production of energy, with low carbon dioxide emissions, where countries such as France, Sweden, the Netherlands and the United Kingdom, depend on nuclear power to keep emissions low and help meet climate change targets.

(Barrett, 2008) states that for too many years Ireland has been opposed to nuclear power, for want of knowledge more than anything else, noting that nuclear energy gives Ireland another option for electricity generation. It is important to recognise all the proposals and alternatives available for electricity generation (Barrett, 2008). (Kelly, 2008) suggests that the question facing Ireland is whether the country needs to generate nuclear power in the Republic of Ireland, or on the island of Ireland, to secure energy supply and meet climate change targets.

3.3. Nuclear Power: International Context

The World Nuclear Association (2007) states that governments are turning increasingly to nuclear power to achieve: (1) national goals of price stability and energy security; and (2) global goals of environmental preservation through reduced carbon emissions. (Yeager, 2003) stated that electricity generated from nuclear power shows prospects as being a reliable, low-cost source of electricity, providing secure and stable cash flows, offering a hedge against fossil fuel prices and supply volatility and potential to safeguard against escalating environmental requirements.

(Landis, 2007) noted that rising gas and oil prices have pushed up the cost of power and brought nuclear power to a position of clear cost advantage, where from an investor's perspective, the time is now right to invest in companies with existing nuclear power plants. The (Sustainable Development Commission, 2006) believes that nuclear power competes largely with fossil fuel power; and with the price of fossil fuel, especially oil and natural gas, rising, the competitive position of nuclear power would continue to improve over the next decade. According to the (Royal Academy of Engineering, 2004), the European Union will become increasingly dependent on imported natural gas for energy, estimating that by 2030; almost 34% of all energy requirements will be from imported natural gas. The (Royal Academy of Engineering, 2004), believes that nuclear power is an established technology that can meet base-load energy requirements, with zero carbon emissions. The (World Nuclear Association, 2007), in addressing the economics of nuclear power, states that nuclear power is cost competitive with other forms of electricity generation, except where there is direct access to low-cost fossil fuel. According to the (International Energy Agency, 2005), nuclear power had increased in competitiveness over the previous seven years (since 1998), mainly on the back of rising natural gas prices and higher capacity factors for nuclear plants, with the study stating that findings did not factor in the cost of carbon when calculating the cost of power from natural gas and coal fired plants. The cost of generating power from a nuclear plant was on a par, if not cheaper, than natural gas when compared across 12 countries (International Energy Agency, 2005).

The (Institute for Energy and Environmental Research, 2006), argues that if one takes into account the greenhouse gas emissions associated with uranium mining, transport, processing, storage and disposal, that nuclear energy has advantages over traditional energy sources like coal and natural gas and is environmentally competitive with alternatives like wind power. (Duetch *et al.*, 2003), states that nuclear power looks more competitive when the cost of CO₂ emissions are taken into account when compared with fossil fuel plants, whilst achieving global carbon emissions stabilisation and reduction targets. The (International Energy Agency, 2007) notes that the international community is looking to technology to help meet the pressing challenges of energy security, environmental protection, economic growth, and the need for clean energy technologies. Security of energy supply has entered the political debating arena, amid concerns about the ability of both gas and electricity systems to meet peak demands and the realisation that gas and oil self-sufficiency are coming to an end, combined with concerns about the long-term reliability of major overseas sources of supply (Sustainable Development Commission, 2006). The (Department of Communications Energy and Natural Resources, 2007), believes that Ireland requires robust electricity generation capacity to ensure consistent supply to consumers and all sectors of the economy. According to the (Department of the Environment, Heritage and Local Government, 2007), there is scientific consensus that global warming is happening, and an economic consensus that the costs of inaction will greatly outweigh the costs of action, and that progressive climate change policies must be based on innovation and investment in low-carbon technology.

(Deutch *et al.*, 2003) states that studies into the future use of nuclear power is necessary to face the challenges of future energy needs without emitting carbon dioxide (CO₂) and other atmospheric pollutants. According to the (President's Committee Of Advisors On Science And Technology, 2008), a possibility exists for the recovery of uranium from sea water, and with a potential recoverable quantity of some 4 billion tonnes, this could support over 3,000 GW of installed nuclear capacity for 6,500 years. According to the (International Energy Association, 2006), competitive costs and CO₂ mitigation potential are not the only precondition for nuclear power's expansion. Public acceptance, final waste management and proliferation risk are important areas to further address and develop. If not addressed, nuclear power is unlikely to expand and its share in electricity generation might be dropping in the future (International Energy Association, 2006). According to (Cummins, 2006), through the ban on nuclear power and the refusal to allow debate on the possible use of nuclear power to generate electricity, the Irish government has turned its back on a technology that can help maintain the Irish economy and tackle climate change.

4. Research Findings

4.1. Main Findings

The main finding of this research study highlights the need for more information on nuclear power to be made available in Ireland. There were 86% of participants, as shown in Figure 1, generally in favour of more information being made available in Ireland on nuclear power. Conversely, only 13% of participants were not in favour of more information being made available.

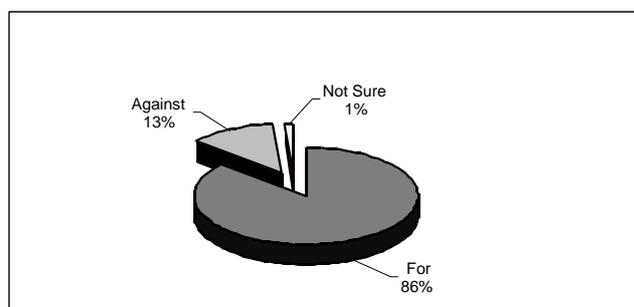


Figure 1: More Information on nuclear power made available

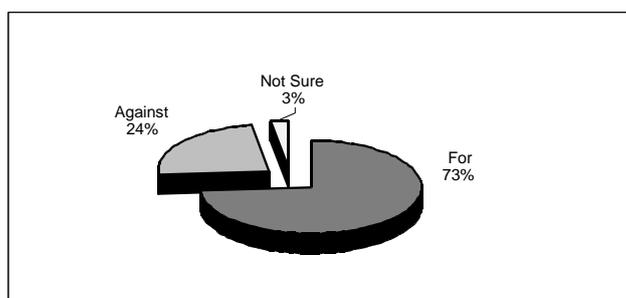


Figure 2: National debate on nuclear power

More information on nuclear power being made available in Ireland is supported by 86% of participants (Figure 1), and 73% (Figure 2), are in favour of a national debate on the use of nuclear power in Ireland. (Barrett, 2008) suggests that nuclear plants are a mystery to a high percentage of the Irish public, and if this issue could be addressed, then an objective debate could take place on the use of one system, rather than another, for the production of energy.

(Barrett, 2008) went further to state that the element of fear on nuclear power needs to be removed from the debate to obtain a level playing field so that nuclear power can be debated along with other methods of generating electricity. (McWilliams, 2005) called for nuclear power to be revisited as a potential logical alternative to fossil fuels, given the depletion of the world's fossil fuel resources and the fact that carbon emissions are unsustainable, nuclear power is a logical alternative.

(McWilliams, 2006) also suggests that Irish people should open their minds to the possibility that nuclear power is part of the energy solution, not part of the problem. According to (Quinn, 2009), Ireland is unable to consider nuclear power plants because of official State opposition to it, with a prohibition on nuclear power written into law as part of the Electricity Regulation Act of 1999, precluding even the consideration of nuclear power.

This current research has found that attitudes and opinions on nuclear power vary depending on the nature of the issues being discussed. This is evident from the findings where, initially, a large proportion of participants are not in favour of building a nuclear power plant in Ireland, but when introduced to concepts such as lower electricity costs, enhanced security of supply and lower carbon dioxide emissions, the participants were marginally in favour of building a nuclear power plant in Ireland.

This current research has found that 58% of participants, as shown in Figure 3, were generally against building a nuclear power plant in Ireland, with only 29% in favour.

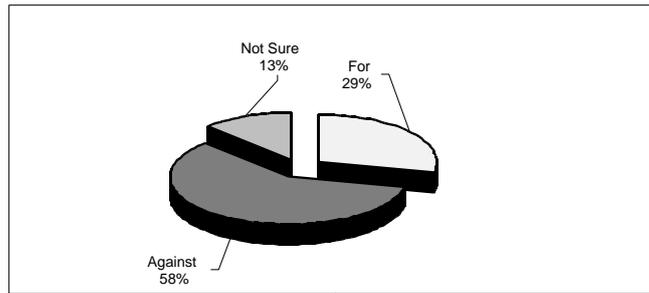


Figure 3: Build a nuclear power plant in Ireland

This research also found that when introducing the concept of lowering electricity costs through the use of nuclear power, the findings presented a notable change of attitude, where 49% of participants, illustrated in Figure 4, were not in favour of building a nuclear plant. This finding represents a decrease of 9% (58% to 49%) in participants who are not in favour of building a nuclear power plant in Ireland. Conversely, there was increase of 9% (29% to 38%) of participants who are in favour of building a nuclear power plant.

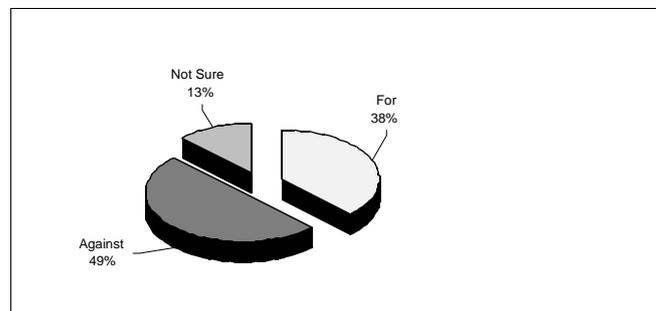


Figure 4: Build nuclear + lower electricity costs

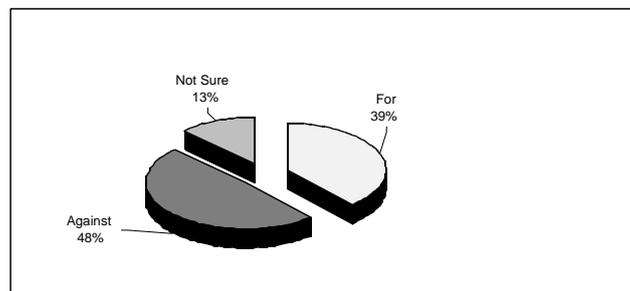


Figure 5: Build nuclear + lower electricity costs + enhanced security + enhanced security of supply

This research found that when the concept of enhanced security of supply was introduced in parallel with the previously mentioned concept of lower electricity costs, there was only a 1% drop (49% to 48%) in participants who are against nuclear power, and no drop in support for nuclear power, as shown in Figure 5.

According to the (World Nuclear Association, 2007), while there is international recognition of the role that nuclear power has played in satisfying various policy objectives, including energy security of supply, reducing import dependence and reducing greenhouse gas or polluting emissions, nuclear power must demonstrate its economic competitiveness as well as its life cycle advantages. (Hutton, 2008) states that nuclear power has provided the UK with safe and secure supplies of electricity for more than half a century, being one of the few proven low-carbon technologies that can provide base load electricity. (McWilliams, 2006) and (Quinn, 2009) both argue that Ireland cannot assess the merits or otherwise of nuclear power due to the current prohibition on nuclear power. A further significant finding in this research study occurred when the concept of lower carbon dioxide emissions was presented with the previously mentioned concepts of lower electricity costs and enhanced security of supply.

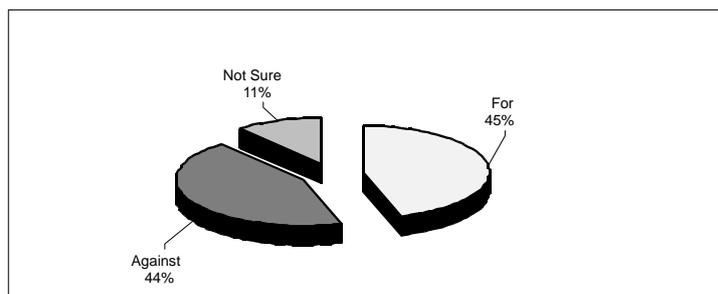


Figure 6: Build nuclear + lower electricity costs + enhanced security of supply + lower carbon dioxide emissions

In Figure 6, the findings show that 45% of participants are generally in favour of building a nuclear power plant in Ireland, and 44% against, when the three concepts of lower electricity costs, enhanced security of supply and lower carbon dioxide emissions were introduced together. This finding illustrates a 14% decrease in those participants (58% to 44%) who are not in favour of building a nuclear power plant in Ireland. The findings also show a 16% increase (29% to 45%) of participants who are in favour of building a nuclear power plant in Ireland. This research has also found that participants are generally not in favour of nuclear power being the dominant method for generating electricity in Ireland. The survey revealed that 56% of participants, as shown in Figure 7, are not in favour of nuclear power being the dominant method for generating electricity, with 31% of participants in favour of the proposition.

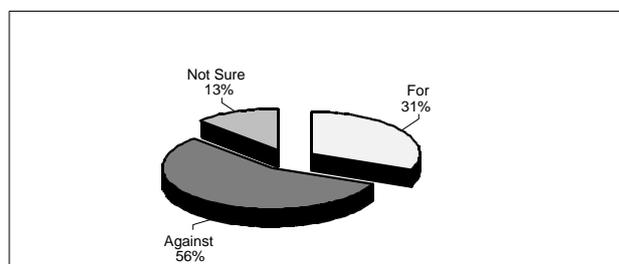


Figure 7: Nuclear power as the dominant method for generating electricity

This research has also found that support exists for a feasibility study to be undertaken to investigate the use of nuclear power in Ireland, with 59% of participants believing such a feasibility study should be undertaken, illustrated in Figure 8.

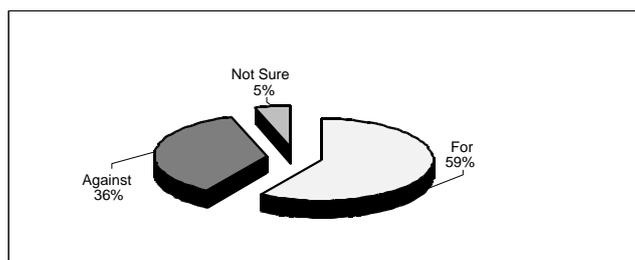


Figure 8: Nuclear power feasibility study

Notably, on the issue of modern nuclear power technology and improved safety, 50% of participants would be in favour of building a nuclear power plant in Ireland, illustrated in Figure 9, if it could be proven that the technology was safer, with 34% of participants not in favour.

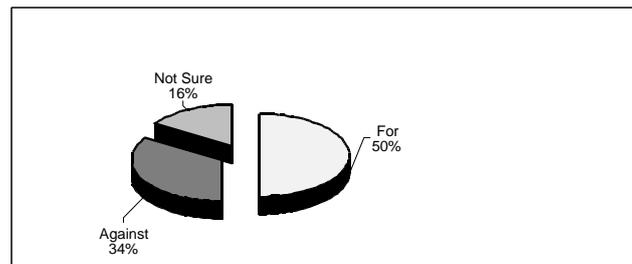


Figure 9: Build nuclear power – safer technology

According to (Deutch *et al.*, 2003), U.S. public attitudes on nuclear power are informed almost entirely by their perceptions of the technology, where views on nuclear waste, safety, and costs are critical to their judgments about the future deployment of this technology. Technological improvements that lower costs and improve safety and waste problems can increase public support substantially (Deutch *et al.*, 2003).

5. Recommendations and Conclusions

5.1. Recommendations

5.1.1. Feasibility Study on Nuclear Power

In light of the findings of this current research study, it is recommended that the Irish government commission a feasibility study into the use of nuclear power in Ireland for generating electricity in the future. This feasibility study should investigate all aspects of nuclear power, especially, the economics of nuclear power, environmental issues, security of supply, safety concerns, waste management, and non-proliferation.

The feasibility study should be carried out by an outside independent organisation, such as the Organisation for Economic Cooperation and Development (OECD), or the European Commission, or other external entity that has the experience and background to undertake such a study, while maintaining a position of impartiality. The feasibility study should look at both the ‘pros’ and ‘cons’ of nuclear power and the possibility of providing information on nuclear power in Ireland. 59% of participants in this current research for example, are in favour of such a feasibility study, while 65% of participants believe that nuclear power is a relatively cheap source of electricity, and 73% of participants believe that nuclear power is an environmentally friendly method of generating electricity.

5.1.2. Public Information on Nuclear Power

It is clear from the research findings that more information on nuclear power should be provided to the Irish public. Such information should be fair, impartial and based on actual international studies and experiences. Without additional information, it would be difficult to sustain any national debate on the issue of nuclear power and its potential use in Ireland. The earlier recommended government commissioned feasibility study into nuclear power would ideally be a corner stone of all publicly available information.

There is a long history of operating nuclear power plants across Europe and North America from where much information can be obtained, numerous studies and reports published on practically every aspect of nuclear power from operations, costs, technology, fuel (Uranium), safety incidents and safety records, environment benefits and non-proliferation. This research has found clear support for more information on nuclear power to be made available to the public, with 86% of participants supporting more information being made available in Ireland on nuclear power.

5.1.3. National debate on Energy Policy and Nuclear Power

The findings suggest that a broad discussion or debate is required on the potential use of nuclear power in Ireland. This however, should form part of a larger debate on Ireland’s overall energy policy for the future. Issues such as the economics of current energy policy, the status of environmental targets and future security of supply issues should be debated.

In particular, Ireland’s dependency on imported fossil fuels should be rigorously debated and a plan for how the country will generate electricity in the event that fossil fuel supplies are exhausted, as is predicted. This recommendation is supported by these research findings which show solid support, 73%, for a national debate on the use of nuclear power.

This research also reports findings of considerable support for varying the fuel mix in Ireland to ensure a reliable and secure supply of electricity, but also to reduce carbon dioxide emissions through a reduction in fossil fuel usage and an increase in renewable energy.

The (Royal Academy of Engineering, 2004) state that nuclear power has the lowest lifecycle carbon dioxide emissions of all the current electricity generating fuels and technologies. According to (Deutch, 2009), reliable and secure electricity is possible using nuclear power due to the fact that uranium resources, the nuclear fuel, are rising faster than consumption, with current resources sufficient to fuel the deployment of 1000 reactors over the next half century.

5.2. Conclusions

This current research concludes that nuclear power merits investigation at a minimum, to provide further information for decision making. (Yeager, 2003) states that nuclear power shows prospects as being a reliable, low-cost source of electricity, offering protection against supply volatility and potential to safeguard against escalating environmental requirements. (Aylward, 2008) believes that nuclear power is one option which should be debated openly, where Ireland has the facts on nuclear power, and the nuclear option for generating electricity openly considered. According to (Fitzpatrick, 2008), the Irish government should establish a highly qualified expert group to investigate nuclear power, believing that for too many years, Ireland has been opposed to nuclear power, for want of knowledge more than anything else.

An investigation or feasibility study into nuclear power in Ireland does not mean that the government or, the public, support this particular technology, but it is only fair, that Ireland should be able to debate the matter. It is most unjust that such a prohibition policy is currently in existence. There is evidence of concern for the environment and support for a reduction in carbon dioxide emissions. There is also evidence that renewable energy should play a larger role in Ireland's future energy strategy. Certainly, there is evidence of concern on issues such as nuclear safety and nuclear waste storage. Interestingly, there are clear signals from this research however, that nuclear power is perceived to contain some tangible benefits that may in the future warrant a review of the prohibition on its use in Ireland.

6. References

- Ahern, B. (2007). *Delivering a Sustainable Energy Future for Ireland*. Dublin: Department of Communications Energy and Natural Resources. White Paper.
- Aylward, B. (2008). "Joint Committee on Climate Change and Energy Security". *Session with the Better Environment with Nuclear Energy (BENE)*, 29th October, 2008. <http://debates.oireachtas.ie/DDebate.aspx?F=CLJ20081015.xml&Node=H3&Page=5>
- Barrett, S. (2008). "Joint Committee on Climate Change and Energy Security". *Session with the Better Environment with Nuclear Energy (BENE)*, 29th October, 2008. <http://debates.oireachtas.ie/DDebate.aspx?F=CLJ20081015.xml&Node=H3&Page=5>
- Campbell, C. J. (2000). "Peak Oil", *Presentation at the Technical University of Clausthal*. Available at: <http://energycrisis.org/de/lecture.html>.
- Cummins, M. (2006). *Kyoto Compliance – Review of the national climate change strategy*. *Better Environment with Nuclear Technology (BENE)*.
- Department of Communications Energy and Natural Resources (2007). *Delivering a Sustainable Energy Future for Ireland*. White Paper.
- Department of the Environment, Heritage and Local Government. (2007). *National Climate Change Strategy 2007 – 2012*. Dublin: DEHLG. Available at: www.environ.ie
- Department of Trade and Industry (2003), *Our Energy Future – Creating A Low Carbon Economy*. UK Department of Trade and Industry.
- Deutch J., E. Moniz, S. Ansolabehere, M. Driscoll, P. Gray, J. Holdren, P. Joskow, R. Lester, and N. Todreas. (2003), "The Future Of Nuclear Power", Available at: <http://web.mit.edu/nuclearpower>
- Deutch, J. (2009), "Update of the MIT 2003 Future of Nuclear Power Study", MIT Press. Available at: <http://web.mit.edu/nuclearpower/pdf/nuclearpower-update2009.pdf>. Retrieved 2009-05-18
- Doran. (2009), "Collective Energy Security: A New Approach for Europe", Center for European Policy Analysis. Available at: http://www.ensec.org/index.php?option=com_content&view=article&id=177:collective-energy-security-a-new-approach-for-europe&catid=92:issuecontent&Itemid=341
- Energy Information Administration. (2009), "International Energy Outlook 2009", U.S. Department of Energy. Available at: [www.eia.doe.gov/oiaf/ieo/pdf/0484\(2009\).pdf](http://www.eia.doe.gov/oiaf/ieo/pdf/0484(2009).pdf)

Managing the Future Energy Policy for Ireland: Examining the Role of Nuclear Power

- Fitzpatrick, M. (2008) "Joint Committee on Climate Change and Energy Security", *Session with the Better Environment with Nuclear Energy (BENE)*, 29th October, 2008. <http://debates.oireachtas.ie/DDebate.aspx?F=CLJ20081015.xml&Node=H3&Page=5>
- Gallis, P. (2009), "NATO and Energy Security", European Affairs Foreign Affairs, Defense, and Trade Division. Congressional Research Service. Available at: http://www.usembassy.at/en/download/pdf/nato_energy.pdf
- Gittus, J. (2007) *A Future for Nuclear Energy in Ireland?*, Royal Academy of Engineering Professor of Integrated Business Development.
- Hutton, M. J.. (2008), "UK Secretary of State for Business, Enterprise and Regulatory Reform". Press release. Available at: <http://environment.uk.msn.com/climate-change/article.aspx?cp-documentid=7354928>
- International Energy Agency (2005), "Projected costs of generating electricity, 2005 update", Available at: <http://www.iea.org/textbase/nppdf/free/2005/ElecCost.pdf>
- International Energy Agency (2006), "Energy Technology Perspective", Available at: <http://www.iea.org>
- International Energy Agency (2007), "World Energy Investment Outlook", Available at: <http://www.iea.org>
- Institute for Energy and Environmental Research (2006), Available at: <http://www.ieer.org/reports/npd.html>.
- Keenan, B. (2008), "The debate on Irish nuclear power should be fuelled by pragmatism", *Independent Newspaper*, Sunday January 13 2008. Available at: <http://www.independent.ie/opinion/columnists/brendan-keenan/the-debate-on-irish-nuclear-power-should-be-fuelled-by-pragmatism-1264931.html>
- Kelly, M. (2008), "Joint Committee on Climate Change and Energy Security". *Session with the Environmental Protection Agency (Dr. Mary Kelly, Chairperson, EPA)*. Available at <http://debates.oireachtas.ie>
- King, S. (2009), *The Irish Examiner* Wednesday, September 09, 2009
- Landis, D. (2007), "Nuclear Ambitions", *Kiplinger's Personal Finance*, February 2007 Edition. Kiplinger Washington Editors Inc.
- McManus, L. (2007), "Joint Committee on Climate Change and Energy Security", *Session with Deputy Eamon Ryan, Minister for Communications, Energy and Natural Resources*, Available at <http://debates.oireachtas.ie>
- McWilliams, D. (2005), "We need to go nuclear". Available at: <http://www.davidmcwilliams.ie/2005/09/12/we-need-to-go-nuclear>
- McWilliams, D. (2006), "Why is it an article of faith to be against nuclear power", available <http://www.davidmcwilliams.ie/2006/11/04/why-is-it-an-article-of-faith-to-be-against-nuclear-power>
- President's Committee Of Advisors On Science And Technology, (2008), "Cooperation In Energy Research, Development, Demonstration, And Deployment". Washington, DC, pp. 5-26 - 5-27. Available at: <http://www.whitehouse.gov/WH/EOP/OSTP/html/P2E.pdf>
- Quinn, Senator Fergal (2009), *Press Release*. Available at: <http://www.feargalquinn.ie/index.php/News/Publications/Fuel-policy-is-fossilised-we-must-consider-nuclear-option>
- Royal Academy of Engineering (2004), *The Cost of Generating Electricity*. London.
- Ryan, E. (2007), "Parliamentary Debates"; *Dail Eireann. 5th December 2007*. Written Answers From Deputy Eamon Ryan, Minister for Communications, Energy and Natural Resources. Available at <http://debates.oireachtas.ie/Xml/30/DAL20071205.PDF>
- Sustainable Development Commission, (2006) *The role of nuclear power in a low carbon economy*. Paper 4: "The Economics of Nuclear Power". London: NERA Press. Available at www.nera.com
- Sustainable Energy Ireland, (2007), *Security of Supply in Ireland: 2007 Report*, Available at: www.sei.ie
- Turvey, F. White, P. Walton, P. Stafford, J. O'Flaherty, T. Duff, D. 2008. "Joint Committee on Climate Change and Energy Security", *Session with the Better Environment with Nuclear Energy (BENE)*, 29th October, 2008. <http://debates.oireachtas.ie/DDebate.aspx?F=CLJ20081015.xml&Node=H3&Page=5>
- World Nuclear Association, (2007), "The new economics of Nuclear Power", London. World Nuclear Association Press. Available at www.world-nuclear.org.
- Yeager, K. (2003), "Nuclear Power: Economic Alternative for Base-load Electricity" Electric Power Research Institute. Available at: <http://my.epri.com>.

Electric Vehicles in Ireland: The Future ?

N. O'Mahony, A. Wright

Cork Institute of Technology, Cork, Ireland

Abstract

This study explores the merits of the Electric Vehicle (EV) as an alternative mode of transportation in Ireland. The EV has been produced for many decades, but never in a mass production capacity. EVs do not require fossil fuels to power them. With fossil fuels being in short supply and as global climate change is deteriorating each year; the EV will help reduce Ireland's dependence on depleting fossil fuel resources. This current study seeks to establish if the EV can be part of the solution to prevent further un-repairable damage to the environment. The findings of this current study, demonstrate that EVs have the potential to contribute to significant CO₂ reductions in the transport sector. This research also outlines the main advantages and disadvantages of the EV, and reviews the future development of EVs. This research study focuses on the technology development, infrastructure development, consumer acceptance and appropriate measures that will facilitate a more rapid utilisation of EV technology. The Irish Government have an existing strategy for the adoption of electric cars, and if successful, the commercial vehicle industry and large trucks may factor into the longer term electric transport strategy. Ireland has some of the best natural resources and the EV will allow Ireland the opportunity to use these resources to generate electricity. This current study seeks to establish if the EV can help Ireland build a successful renewable future for all. The EV is exciting to drive, efficient, has great performance and above all, is environmentally friendly with Zero CO₂ emissions.

Keywords

Electric vehicles, Environmentally friendly, Climate change, Emissions, Government incentives.

1. Introduction

This paper presents an overview of the future of EVs in Ireland. Substituting EVs for the traditional Internal Combustion Engine (ICE) vehicles could help to substantially reduce local environmental pollution and greenhouse emissions, currently emitted from the transport sector. According to (Matthews, 2009), Ireland imports over 90 per cent of its national fuel requirements and, like many other countries, is seeking ways to secure its energy supply. The (Department of Communications, Energy and Natural Resources, 2010) purports that it is the Irish Government's intention to have 10 per cent of all Irish cars powered by electrical methods by 2020. If this policy succeeds, Ireland will be amongst the first in the world with the national infrastructure to support this policy and will be viewed as a global leader in the green economy. Future mobility is now challenged due to the fossil fuel shortages and global warming concerns. The supply of low-priced oil is quickly diminishing, which means that the fossil fuels that previously powered cars, trucks, planes and ships, are becoming scarce and therefore more expensive. (Rose, 2010) states that the recent increases in oil are alarming for the world; and that this situation has to change, and fast. In addition, people are becoming more concerned about the environment, and especially, climate change. The automobile industry is one of the greatest consumers of the fossil fuels which continue to prove a serious problem for the environment and human life. The EV may be of substantial benefit, reducing harmful exhaust emissions in the future.

2. Literature

2.1. EVs: A Renewed Interest

According to (Leitman *et al.*, 1994), EVs were designed to do whatever was required in the past, and can be designed and refined to do whatever is necessary in the future. (Cowan *et al.*, 1996) state that the oil crises in 1973-74 forced many politicians to re-evaluate their transport systems' dependence on an unstable political regime in the Middle East. The oil crises, in particular, encouraged the creation of new EV programmes. Various research studies carried out during the 1970s and 1980s were commissioned, but this research did not result in the creation or recommendation of mass production of EVs. The majority of the research in the 1970s was built on the assumption that the battery technology used in the EV could improve its capabilities quickly. Unfortunately, the improvement in battery technology never came to fruition and electric and hybrid vehicles have therefore remained uncompetitive. The viability of the internal combustion engine vehicle was questioned due to the increase in congestion in the road networks in large cities, the amount of vehicle accidents that claim thousands of lives every year and the air pollution generated. These were some of the main reasons that prompted doubts about the ICE vehicle (Cowan *et al.*, 1996). According to (Dhameja, 2002), by the middle of the 1900's the diminishing oil supplies and the growing demand on the fossil fuels began to re-ignite

the aspirations of alternative energy sources and the discovery of alternate energy solutions. By the mid 1970's oil shortages led to the sudden and aggressive development of EV programs by manufacturers. Shortly afterwards, the concern with the oil supply was reduced as supply stabilised which in conjunction with a slow advancement in the alternate technology for batteries, impeded the development of the EV (Dhameja, 2002).

3. Methodology

Initially, an extensive review of the existing literature was conducted on the automobile industry and current global climate challenges, so as to provide a solid foundation for the research. The literature relays that interest in the EV is on the increase, and that current attention is unlikely to dissipate in the foreseeable future. Reasons range from climate change and environmental pollution, to the gradual depletion of fossil fuels worldwide, and to the nationalistic desire to reduce dependence on imported oil. Each country faces its own unique set of challenges but environmental pressures and the looming background threat of global warming are common to all. For this reason alone, interest in cleaner vehicles as alternatives to traditional gasoline powered cars will grow. EVs stand poised to fulfil a need, to aid the effort to curb harmful emissions from liquid fuels. Technology, economic context, and environmental issues are all aligning to create a more commercial backdrop for EVs.

The purpose of this research is to gain an understanding of the EV, and to determine the future of the EV in Ireland. The methodology is specifically chosen for its appropriateness for this research study. For the purpose of this research, a qualitative approach was taken in the form of semi-structured interviews, as rich informative data was required on the topic. An interview guide emerged from gaps in the existing literature and was used for discussion. A pilot interview was undertaken initially, to establish the validity and reliability of the instrument. The order of the questions was adapted according to the flow and direction of the conversation. The interviewer conducted the personal interviews with twelve purposely picked informed individuals, in a face-to-face context, where the researcher met with the appropriate twelve interviewees and questioned them in detail on EVs in Ireland. These individuals had considerable expertise in the area, and included participants such as the Minister for Transport, the chairman and MD of Ford Ireland, Nissan technical experts, and Green Party members. The data collection process ensured a high level of knowledge and insight into the topic of EVs. Ethical considerations were taken into account, and the data collected and processed guaranteed the responses would only be used for this required study.

4. Findings

This paper analyses the changes that have taken place in the automobile industry in Ireland. This research focuses on how Ireland needs to adapt and prepare for the introduction of the EV. The findings of this current study suggest that to enable the EV to be successful in Ireland, government incentives are required, combined with additional support for car manufacturers to enable manufacturers to migrate the EV into mass production. This will then afford the EV the additional benefit of utilising economics of scale which will help to reduce the final purchase cost of the EV and put the EV in a competitive position.

4.1. The Automobile Sector

The Automobile industry was very successful during the Celtic Tiger era, but since the economic downturn, the motor industry has suffered an unprecedented decline. In 2008, the Government Budget announced that the vehicle registration tax and annual road tax systems were to change with effect from July 1 2008. The new system links the tax rates directly to specific CO₂ emissions (CO₂ g/km) rather than engine size. It established a strong purchasing signal, promoting lower-emissions cars. Participants in this study believe that there is no evidence that the tax change resulted in consumers purchasing smaller cars, but there has been a distinct shift towards diesel-fuelled cars. The share of new diesel engine vehicles has been increasing, from 10 per cent in 2000, to 34 per cent in 2008. During the first half of 2008, the share was 28 per cent of sales; after July, it increased to 54 per cent.

4.2. The Future of the Automobile Sector

The EV will have a role in the future of the automobile industry, because of the need to reduce emissions. It will always be necessary to buy new vehicles, as a vehicle's performance and ability will decline with age. Also, the public transport facilities are very underdeveloped and this allows the opportunity for increased car sales. The future for the automobile sector is looking promising, with the electric vehicle providing a great opportunity, but it is clear from the key interviewees in this study, that there is further research and development required for the successful advancement of the EV. Is the infrastructure and technology in place to embrace the EV? EVs require an electrical charge to operate their motors in order to allow them to operate. The ESB are installing an initial infrastructure to ensure that 3,500 charging points are available throughout Ireland. There is a perception among participants that there will be high costs associated with the EV, and a lot of uncertainty in the market place in relation to the price of the EV. This paper also found that the

EV is still expensive to purchase, and for the next few years or so, the EV will only suit a niche segment of people in urban cities. The driving range is an issue, and until this is resolved people will not fully transfer to this mode of transport. The EV will need to advance to the stage where mass production is a reality; this in turn will help reduce the price by economies of scale, but it will not be until this point that the traditional motorists will choose to drive the EV.

5. Discussion and Recommendations

From this current research it is clear that Ireland has to reduce its dependence on fossil fuels. As the prices of these fuels are continually increasing, there is an increased focus to further investigate alternative transportation. The EV is clearly an option to help reduce Ireland's dependence on foreign imported fossil fuels and to reduce Ireland's GHG emissions.

A major finding from this paper is that there is a lack of consistent information in the public domain about the EV. There is a lot of confusion between motor manufacturers and the consumer. The Government along with the automobile manufacturers need to outline a clear communication strategy to educate consumer about the EV. It is clear from this current research that there is increasing interest among customers for EVs. Manufacturers in conjunction with the government need to respond by stepping up efforts to bring these models to the marketplace alongside some of the latest-generation, fuel-efficient petrol and diesel powered models. If, for example, a traffic restriction zone comes into effect in Dublin, the only option might be the EV in these restricted zones. The researcher recommends that the internal combustion engine manufacturers consider the environment in their future strategic decisions and management strategies. It is clear that the car manufacturers have certainly made and continue to make more fuel efficient and lower emission vehicles.

The Government envisage that 10 per cent of all cars will be electric by 2020 in Ireland. It is intended that the EV will replace the internal combustion engine during the next two to three decades. This transport evolution will occur chiefly because of the decline in world oil production. Electricity from renewable energy and other sources will be the most feasible alternative fuel. These societal benefits come at high cost to the owner of EVs in terms of price, driving range, availability, loading capacity, speed and acceleration. The practicality of the EV is unfortunately hampered by the lack of infrastructure for recharging. Skilful marketing will certainly be needed if EVs are to be accepted and diffused throughout Ireland.

5.1. The Future of the Automobile Industry

The future of the automobile industry will most certainly focus on reducing the CO₂ emissions from this sector. "The European union is mandating maximum levels of CO₂ emissions from vehicles". This mandating of more efficient cars will ensure that car manufacturers provide more fuel efficient vehicles and more environmentally friendly. "They are going to push the manufacturers into producing cleaner cars and more recyclable cars and clearly the mix of units that is going to be sold will be an amalgam of the conventional vehicles and most certainly, the EVs". While the EV has some issues, the Hybrid vehicle will greatly assist with the transition from the internal combustion engine to the EV. Another significant finding that emerged from this research is that the Hybrids are more suited to the heavy commercials. Until the EV has been fully developed the Hybrids will help the EV get there in the mean time".

5.2. The Role of the EV

According to (Leitman *et al.*, 1994), the EV can solve a lot of transport concerns. EVs bypass high energy prices, they are relatively low cost to charge and have zero exhaust emissions. EV predominantly use electricity from power plants to charge their batteries, they can also use electricity that has been developed from solar, wind and any other renewable resources. If all EVs used electricity generated at power plants to recharge their batteries, the amount of emissions generated to produce this electricity would still be less than the emission emitted by a similar amount of vehicles propelled by the internal combustion engine. This current study also finds that a major factor in the future of the EV will be its ability to equally match and in certain cases exceed the capability and practicality of the internal combustion engine vehicle. The EV will need to be aesthetically pleasing to the point that it should almost be an extension of the person buying the car. The EV will also need to meet and exceed the structural and safety attributes of the current internal combustion engine (Leitman *et al.*, 1994). It is guaranteed that the EV will be subject to a vast array of future improvements to meet customer expectations such as top speed, driving distance range and even increased overall efficiency. "From an environmental point of view the EV is clearly sought after, but it is not without its limitations. Some of these will be difficult to overcome in the short term". "I'm quite sure in the fullness of time that there will be a win, win situation for EVs". The EV is going to challenge current models; there will need to be a revised business model to adhere to the EV.

Previous research by (Brooks, 2010) has shown that perhaps both the Hybrid and the EV technologies will find their own large niches in due course, but it is nevertheless striking that the majority of the world's automobile manufacturers seem to be supporting electric vehicles over plug-in hybrids. The EV is one of the "great white hopes of the industry" it does have a great "buzz about it". A further interviewee stated that the "EV is here to stay but not in any total domination role". Participants believe that "the EV is more of a niche for big cities for —city slickers, for a typical Londoner", but the majority believe that the EV is very desirable and appeals to a large segment. "People from all walks of life from the young to the old have a genuine interest in the EV. It suggests that it is funky and safe for future generations". An important finding is that the automobile industry is continuously evolving; "look at the big picture and back in time, we started with the steam engine, we are now in the Internal Combustion Engine age; alternative fuels are going to be the future for transportation".

5.3. EVs: Infrastructure

To enable and encourage widespread consumer adoption and utilisation of EVs, an infrastructure with sufficient public recharging locations to afford drivers the opportunity to recharge their vehicle on a regular basis during the day will be necessary (Tanaka, 2009). This type of infrastructure will effectively increase the daily driving range of EV and increase its level of practicality and broaden its appeal to the end user. Public charging infrastructure could include opportunities for rapid recharging, either via fast recharge systems or via battery swapping stations that allow the quick replacement of discharged battery packs with fully charged battery packs.

The ESB have committed to install 3,500 charge stations 'juice points' in Ireland. 2,000 of these will be installed in the homes for the first 2,000 consumers to purchase EVs with the remaining 1,500 being installed strategically throughout Ireland. "There is momentum from the ESB and there certainly is a business plan behind it also". Another interviewee felt that the electric infrastructure is not available at the moment, but this is all going to change. Ireland does lack the infrastructure for the EV, but implementing the infrastructure should be possible. "It's just a case of adding the charging points to the existing stations", but charging times and the cost of the EV are important factors. "Two important aspects for us are cost of the EV as well as the charge times". It is necessary to understand the obstacles facing electricity utilities, and the emerging EV industry. It is now essential to develop an understanding of the intermediary technologies that will aid the development of an EV charging infrastructure. It is only through gaining a comprehensive knowledge of the uses of these technologies that an efficient and viable Infrastructure will be established. Coupled with accurate projections of consumer charging behaviour and the subsequent effect on the grid, it is possible to plan the most effective uses of these technologies.

5.4. Challenges for the EV

The major concerns facing the EV industry is the driving range, top speed, and purchase cost of the vehicle (Leitman *et al.*, 1994). Ultimately, the batteries will determine the cost and the performance of EVs. The only waste elements from the EV are its battery. According to (Rose, 2010; & Brooks, 2010), driving range is a crucial factor. Drivers who use their cars only for commuting and driving around town might manage quite easily—if there are sufficient charging points. There are the large numbers of commuters who regularly travel over 100 kilometres a day and this would create further problems—range continues to be an issue. "I would not be impressed about charging a vehicle. It's fine if you are doing short distances. If I was travelling a longer distance and had to stop every 80 km or so to charge it then, I would have an issue with this". Another negative is that the EV is a nice pleasant drive, but there is a safety issue that needs to be addressed due to the lack of drive train noise, "pedestrians will not hear the electric approaching".

It has emerged from this study that the cost of the EV is still expensive. Consumers will not benefit from the reduced savings on fuel savings, as people that drive short distances do not spend a substantial amount of money on fuel. "The cost of the EV is incredible; people are not going to reap the benefits of potential fuel cost savings. People that are driving short distances are not spending the money on fuel compared to those driving longer distances". The EV needs to be attractive. "I think that the EV needs to be more than just an unexciting quiet clean vehicle. The EV needs to be every bit as exciting if not more than the other vehicles". Now may not be an appropriate time to launch the EV with only a few vehicle purchases. "The EV is being brought to market at the wrong time".

5.5. Opportunities for the EV

According to (Leitman *et al.*, 1994), the internal combustion engines are nearly at the end of their product / technological life cycle. As the technology for internal combustion engine declines, technological advances for EVs are only at the embryonic stage. Once lithium battery technology becomes the standard, EVs will be able to travel 300 to 600 miles. Unquestionably, the future looks bright for EVs because the best is yet to come. (Rose, 2010) states that there are several attractive reasons to purchase an EV, you won't be subject to road tax and you have 'done your bit' for the environment, and the more widespread adoption of the EV will help to reduce the cost.

Powered by lithium-ion batteries that can be recharged in just minutes, these nimble commuter cars might have a range of more than 180 miles, depending on the battery pack. For example, an EV with a range of 100 miles would cost less than one with a 200 mile range. Larger EVs probably will be plug-in hybrids that have a gasoline, diesel or an alternative fuel engine that combines the virtues of both electric motors and internal-combustion engines, and burns some form of renewable fuel. Hydrogen or methanol fuel-cell range extenders could follow in a decade or two. Ireland is a perfect location to operate a pilot programme and is continuing to develop the technology and infrastructure required. It is important that Ireland avail of the opportunity on the international stage to gain a competitive edge. Urgent action is required to ensure that Ireland will benefit from first mover advantage. "The EV technology will allow us as a nation to lead other countries". "The only choice that Ireland has to make is, do we want to lead with the EV revolution, or do we want to wait until someone takes the reins"?

5.6. The Future of the EV in Ireland

According to (Dennis, 2010), The Chevrolet Volt, and the Nissan Leaf were presented in select markets in 2010, and tens of thousands of people have their names on waiting lists to reserve one. Mitsubishi and Ford plan to debut their electric cars, and several other automakers are endeavouring to get into the EV industry as soon as possible. (Dennis, 2010) states that electric car advocates believe that as battery technology continues to advance, electric model prices someday will fall to the point that millions of people will be driving an EV. A significant finding supporting (Rose, 2010) is that "ultimately, it will be the price of petrol that will turn people to use electric cars". For now, we all need to be aware that the case for EVs depends on vast, profound changes being made to our lives. "The oil crisis is obviously getting worse. I think that when people get their heads around the electric car, people will gradually begin to drive it". Participants in this research believe that it is important for people to understand that there is "no quick fix".

6. Recommendations for Practice

Ireland has to reduce its dependence on fossil fuels, and there is an increased urgency to further investigate alternative transportation. The EV is clearly an option to help reduce Ireland's dependence on foreign imported fossil fuels and will also help to reduce Ireland's GHG emissions. There are a number of obstacles that EVs must overcome to ensure that they succeed commercially. Successful business models will need to be developed to overcome these following obstacles: The monetary savings achieved by the fact that the EV does not require fuel can be used to offset the battery costs in a manner that may be much more acceptable to consumers. A car with a limited driving range (e.g., 150 km) will need to have ample opportunities to recharge its battery. Recharging stations will be needed at high-traffic locations such as train stations, shopping centres, and public parking areas. Rapid recharge or battery swapping systems may also be important, particularly on highways and along other routes where a quick recharge will be needed. EVs will best suit urban areas, because the technology is still developing and there is clearly an issue with the driving range and high cost. Eventually, mass production will help to reduce the cost of the EV. This study has exposed the fact that there is a lack of consistent information about EVs. The Government along with the automobile manufacturers need to outline a clear communication strategy to educate the consumer.

It is clear that there is increasing interest among customers for EVs. Manufacturers in conjunction with the government need to respond by stepping up efforts to bring these models to the marketplace alongside some of the latest-generation, fuel-efficient, petrol and diesel models. If a traffic restriction zone comes into effect in Dublin, for example, the only option might be driving the EV there.

It is also evident that consumers are very price conscious due to the economic downturn. Therefore, the price of the EV needs to be introduced at an attractive launch price to consumers. It is also clear that people would like to be more environmentally aware. Due to the fact that the Government has set out an objective of 10 per cent of all cars to be run by electricity by 2020 in Ireland, the researcher recommends that the government have adequate support in place to ensure that this objective is reached to ensure the future success of the EV.

It is envisaged that 'EVs' will replace the internal combustion engine vehicles during the next two to three decades. This transport evolution will be driven chiefly by the decline in world oil production. Electricity from renewable energy and other sources will be the most feasible alternative fuel. These societal benefits come at high cost to the owner of EVs in terms of price, driving range, availability, loading capacity, speed and acceleration. In addition, the practicality of the EV is hampered by the lack of infrastructure for recharging. Overall such a product, given current practicalities, is not a very attractive product to potential customers. Beside supportive national policies, skilful marketing is needed to get it accepted and diffused throughout society.

Automobile manufacturers need to demonstrate to people how the EV operates and performs as there is generally a lack of information and confidence about EVs in the market place. Manufacturers should offer demonstration fleets of EVs to afford people the opportunity to drive these vehicles. This would help drivers to become familiar with the EV. A training course would also help to ensure that useful data can be produced under realistic operating conditions.

7. Conclusions

This current research has established that the role of the EV is vital in defining the future of the automobile industry in Ireland. In the past decade, the automotive industry has started to change the way fuel is viewed. A multitude of exciting new technologies emerged, and among them are some very real and practical cars that are increasingly efficient and economical. This paper has examined the key elements of the 'EV in Ireland' as it is perceived now, and the issue of the over dependence on imported fossil fuels for the transport sector. The EV is still a developing technology and does require further development before it is brought to market. The previous negative perceptions of the EVs still resonates in consumer's minds, so there is a task to ensure that consumers understand that this generation of the EV is a reformed desirable product. Transport in Ireland has become unsustainable, primarily as a result of years of poor planning, leading in turn to an over dependence on the private car as a single means of transport. Technological development, infrastructure development, and consumer acceptance with appropriate measures will facilitate a more rapid uptake of EV technology in the future. The automobile industry is now forced to consider vital green and consumer issues and, if EVs are the future, it is fair to say that the future is still one governed by many uncertainties requiring clever executive thinking and forward governmental planning. Other issues and limitations that will need to be seriously examined and researched in the near future, if Ireland is going to take the expansion in EVs seriously, is the situation in relation to the national grid and the vital ability to supply the necessary electricity.

8. References

- Brooks, G. (2010), *Hybrids and EVs*, http://www.just-auto.com/management-briefing/hybrids-and-electric-vehicles-part-3_id105665.aspx?lk=dm, accessed August 29, 2010.
- Cowan, R. Hultén, S. (1996). *Escaping Lock-in: the Case of the EV (Technology Forecasting and Social Change)*, Stockholm:University of Western Ontario and Stockholm School of Economics.
- Dennis, F. (2010), *The future of the electric car*, <http://theweek.com/article/index/206278/the-future-of-the-electric-car>, accessed August 27 2010.
- Dhameja, S. (2002), *EV Battery Systems.*, USA: Butterworth-Heinemann.
- EAAEV. (2008), *Why Electric cars?*, <http://www.alternative-energy-news.info/technology/transportation/electric-cars>, accessed 21 March 2010.
- Ehsani, M., Goa, Y. and Emadi, A. (2009) *Modern Electric, Hybrid Electric and Fuel Cell Vehicles Fundamentals Theory and Design*, Second Edition, US: CRC Press Taylor & Francis Group.
- Leitman, S., Brant, B. (1994) *Build your own EV*, accessed 29th August 2010. http://books.google.ie/books?id=jf4jrFjtH7cC&printsec=frontcover&dq=1st+electric+car&hl=en&ei=8RR-TLiKB8iRswbk5uSVCQ&sa=X&oi=book_result&ct=result&resnum=1&ved=0CC8Q6AEwAA#v=onepage&q=1st%20electric%20car&f=false.
- Matthews, N. (2009), "On the wave of our future", *Business and Finance*, 29 January, p. 44.
- Roberts, G. (2010) *THE WEEK THAT WAS: "Look ma, no petrol"*, http://www.just-auto.com/comment/look-ma-no-petrol_id105279.aspx?lk=dmf. accessed 3 August, 2010.
- Rose, D. (2010), *Has the spark gone out of electric cars?* <http://www.dailymail.co.uk/home/moslive/article-1304123/Has-spark-gone-electric-cars.html#> accessed 27 August 2010.
- Tanaka, N. (2009), *Technology Roadmap Electric and plug-in Hybrid EVs*, International Energy Agency 2009, http://www.iea.org/Papers/2009/EV_PHEV_Roadmap.pdf, accessed 26 June 2010.
- The Department of Communications, Energy and Natural Resources. 2010. *EVs charge pointslaunched*. <http://www.dcenr.gov.ie/Press+Releases/Electric+vehicle+charge+points+launched.htm>, accessed 1 September 2010.

Forecasting an EU farm gate milk price using the system dynamics methodology

D. Bergmann

Department of Mathematics
Cork Institute of Technology, Cork, Ireland
e-mail: dpbergmann@t-online.de

Abstract

In this work an economic model of the EU dairy industry has been developed using the system dynamics methodology. The EU dairy industry is characterized by a rapidly changing policy and trading environment. In response to these changes it is expected that complexity and dynamics in this industry will increase. In contrast to existing models (for example the FAPRI model) which are based on partial equilibrium the system dynamics model explicitly focuses on the use of feedback loops and dynamic cause effect chains.

The system dynamics model captures the interdependencies of the major elements among the supply chain of the dairy market. The milk price in different scenarios has been forecasted and areas for further development outlined.

Keywords

Agriculture, EU Dairy industry, System Dynamics, Forecasting

1. Introduction

In recent times the EU dairy industry is characterized by a rapidly changing policy and trading environment. These changes have contributed to greatly increased farm gate milk price volatility. This volatility has seen the price decrease by 24.4 % from its peak in 2008 to its trough in 2009. Thus forecasting the milk price as well as identifying which elements impacts price become an important component of income planning for farmers and the dairy industry.

In this paper the system dynamics methodology will be used to forecast milk price. System dynamics is an approach for analyzing complex systems. The system dynamics methodology differs from other well known models such as the FAPRI/CARD model (FAPRI, 2011) which is static and based on partial equilibrium as system dynamics focuses on the use of feedback loops and the dynamics which can arise from time delays.

System dynamics has been applied to a wide range of problems including using it to analyze the coffee market (Osorio and Aramburo, 2009) or the oil industry (Morecroft, 2007). However adopting system dynamics to the dairy industry is new and could lead to new insights what factors impact milk price.

In this paper the dairy industry is modelled at an EU aggregate level. This entails that in the model price and costs are the same in each member state which differs from reality where huge differences in price and costs among member states exist.

This paper is organized as follows. In section 2 the system dynamics model will be presented. Section 3 shows the price simulation results based on a number of different scenarios. In Section 4 conclusions are drawn and future improvements to the model are discussed.

2. The system dynamics model structure

2.1. General structure

This section develops the general model structure of the EU dairy system dynamics model. It is based on the generic commodity model developed by Meadows, 1970 and refined by Sterman, 2000. For better presentation the model is divided into five components. Figure 1 summarizes this different components and how they are related with each other (a positive labelled arrow means that the two nodes change in the same direction, e.g. if demand rises price rises, too and a negative labelled arrow means that an increase in one variable will have the opposite effect for the other variable, e.g. if price rises demand will fall). The general structure of the model shown in figure 1 is based on economic theory (Mankiw and Taylor, 2008) which suggests price is set by the interaction of demand and supply. In contrast to standard

models which assumes equilibrium price is reached if the supply curve intersects the demand curve the model in this paper assumes that price information are incorporated into actual supply with a delay. The rationale behind this is that it takes time for farmers to adjust their breeding herd and so the milk supply.

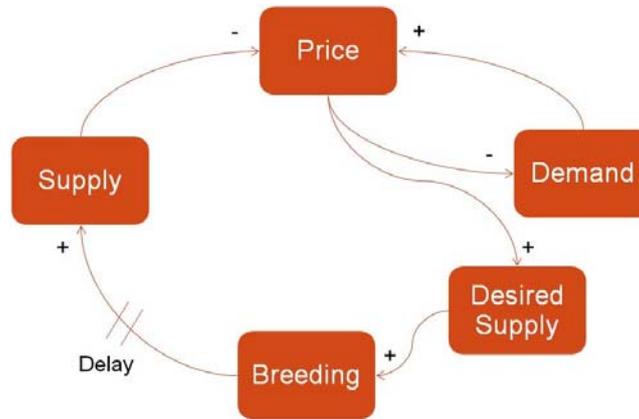


Figure 1: General model structure

2.2. Demand

In this section the demand part of the system dynamics model is presented in more detail. The demand part modelling follows Sterman, 2000 assuming a linear demand curve around reference price and reference demand.

The model is implemented into Anylogic (Xjtechnology Corporation Web Site, 2012) a professional simulation software. The implementation of the demand part is shown in figure 2 (see Appendix A for a description of the symbols). As already mentioned indicated demand depends linearly on price around reference price and reference demand. Because consumers adjust slowly to new price information indicated demand is exponentially smoothed (Sterman, 2000) to give actual demand.

It should be noted here, that demand includes the milk demand by dairy companies to further process it to cheese, butter and other dairy commodities.

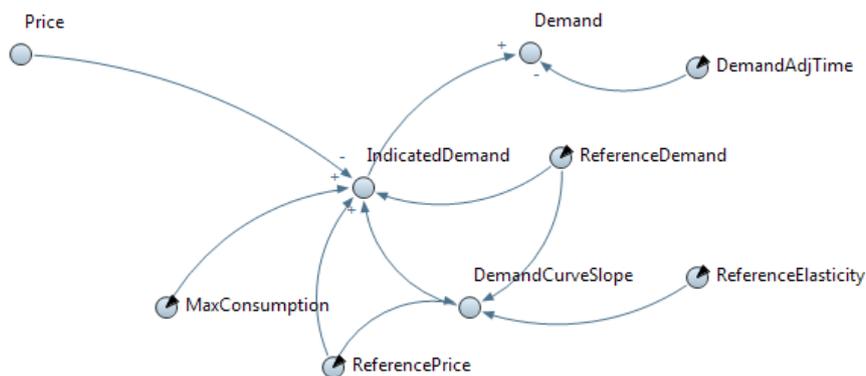


Figure 2: Demand part implementation

2.3. Desired Supply

The general idea of setting future milk supply is that supply is based on expected profitability measured as expected price per tonne divided by expected cost per tonne. This means if the profitability ratio is high new farmers will enter the industry and increase supply whereas if profitability is low some farmers will suffer losses and leave the industry thus lowering supply. In figure 3 is shown how this general idea is implemented into the system dynamics framework. It should be noted that an aggregated economy is modelled.

Future or desired milk supply is set by adjusting current supply upwards if expected profitability is above the reference point and downwards if profitability is below the reference point. To prevent for overreaction supply is exponentially smoothed (Sterman, 2000) over time. From desired milk supply the size of the herd can be derived. It is further assumed that expected costs and expected price is current cost and a smoothed average of recent prices respectively.

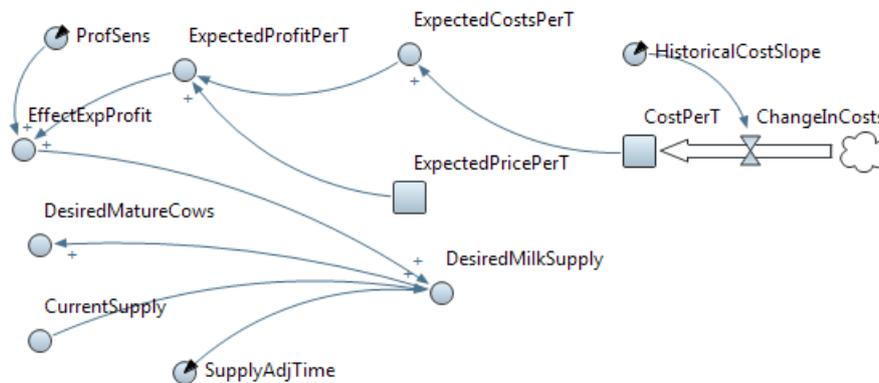


Figure 3: Desired Supply

2.4. Breeding

Figure 4 shows how the actual herd size is adjusted to desired herd size. The difficulty here lies in the fact that it takes time for cows to grow resulting in a delay between the decision of adjusting the herd size and the actual outcome. Here it is assumed that cows pass three different states (from gestation to calf to mature) until they can produce milk. It should be noted that in contrast to expanding herd size cutting it is instantaneous by slaughtering the cows.

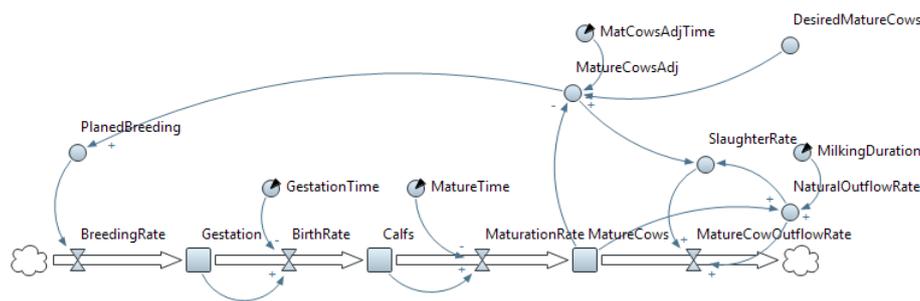


Figure 4: Breeding

2.5. Supply

Modelling actual supply is the most straight forward part because it is just the size of the milk cow herd times the output per cow (also called milk yield). Due to technological advances (e.g. improved milking plants, improved diet and genetics) milk yield increases over time. The milk yield here represents an EU wide average. Milk yields in some member states can deviate substantial from this average. The supply part is presented in figure 5.

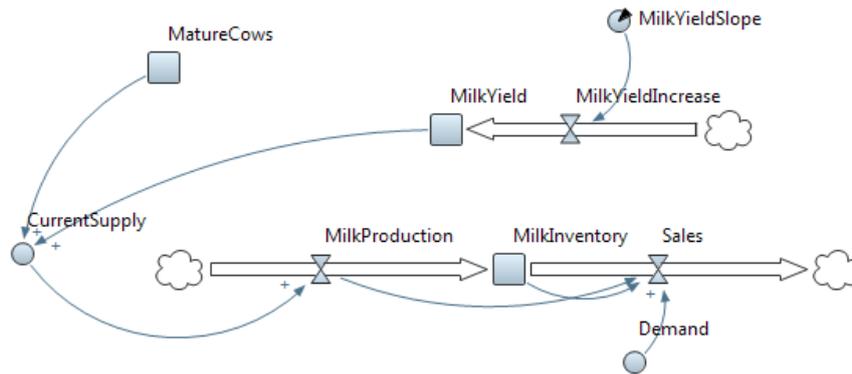


Figure 5: Production

2.6. Price

Modelling the price setting part is a more difficult task. Following Sterman, 2000 a hill climbing heuristic for setting price is used. The hill climbing heuristic is shown in figure 6 which means price is set based on an average of smoothed historical prices adjusted upwards if the supply to demand ratio is below 1 or downwards if the ratio is above 1.

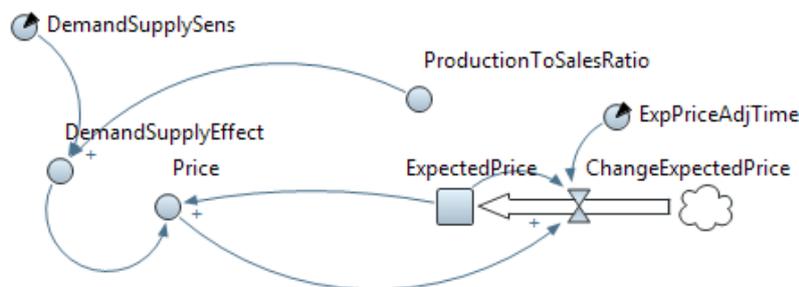


Figure 6: Price Setting

3. Simulation Results

The model was calibrated to yearly EU-15 data from the Farm Accountancy Data Network (FADN) for milk specialized farms from 2000 to 2009. It should be noted that FADN data has some drawbacks. For example costs are calculated for the whole farm and were not allocated to different products which could lead to mistakes estimating costs for mixed farms (see EU Commission, 2010 for a discussion of the FADN methodology). The simulation horizon is 2020. Three different scenarios have been simulated. For the first scenario which is called the base scenario the parameters were estimated using the Anylogic parameter variation method minimizing the mean squared error between historical prices and simulated prices. In the other two scenarios different assumptions about production costs and demand in comparison to the base scenario are used.

3.1. Base scenario

In this scenario it is assumed that costs and milk yield increase linear in time according to their historical trend. The simulation results for the base scenario are summarized in figure 7. The blue curve shows the simulation output whereas the red circles represent the observed price for this period. It can be seen that the model fits the data at the beginning quite well while it could not adequately capture the sharp price fall in 2009.

The simulation results for 2010 till 2020 show a general milk price increase in comparison to historical levels. This observation is in line with the general expectation of rising prices of agriculture commodities due to increasing energy and feed costs (see FAPRI, 2011). It should also be noted that the price oscillates which is a result of the delay between farmers decision to increase output und the time the increase takes effect.

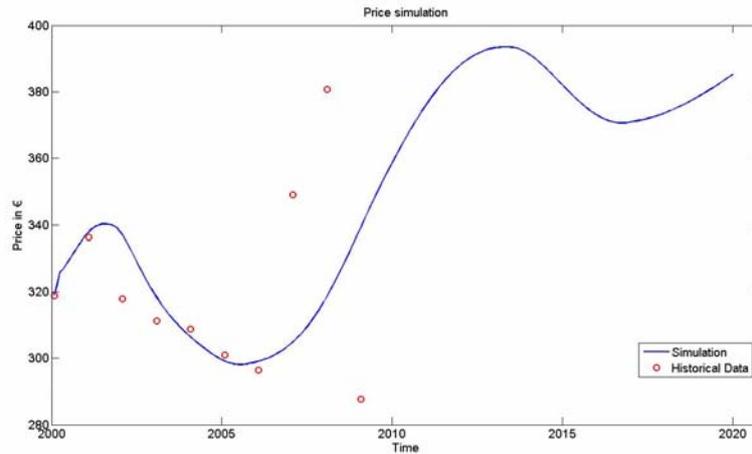


Figure 7: Base scenario simulation

3.2. Increased demand on top of the base scenario

In this scenario a 10 % unexpected and permanent increase in demand versus the base scenario after 10 years is assumed. The simulation results (red line) for this scenario are shown in Figure 8. As general economic theory suggest increased demand will raise price above the price levels of the base scenario. But after about 5 years falls below the price levels of the base scenario although demand is above the demand level of the base scenario. A possible explanation for this unexpected behaviour could be that the sharp price appreciation caused by the demand increase results in an overreaction of farmers to raise supply above a level which would be justified by the increased demand.

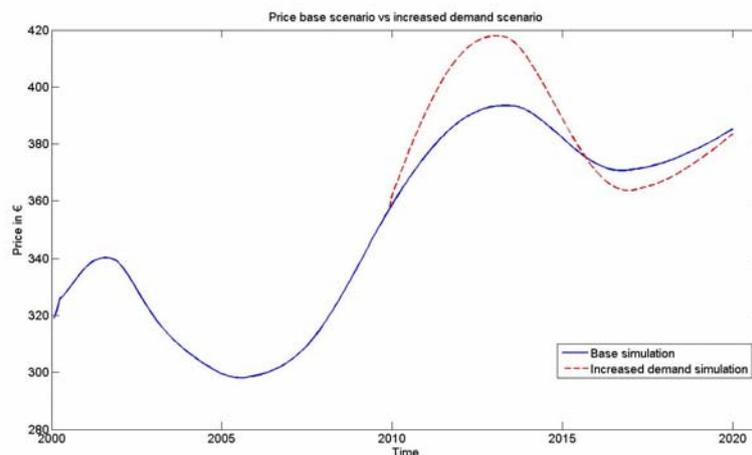


Figure 8: Increased demand scenario

3.3. Increased costs scenario

In this scenario a 10 % unexpected and sustainable increase in milk production costs (e.g. energy or feed costs) is assumed. The simulation results (red line) for this scenario are shown in Figure 9. It can be seen that price in this scenario is well above price in the base scenario. Beside the higher price level the general shape of the curve is comparable to the base case scenario. A possible explanation for this behaviour is that due to the falling profitability some farmers will go out of business resulting in declining supply causing rising price. As unprofitable farmers have left price behaviour is comparable to the base scenario on a higher price level and a lower supply level.

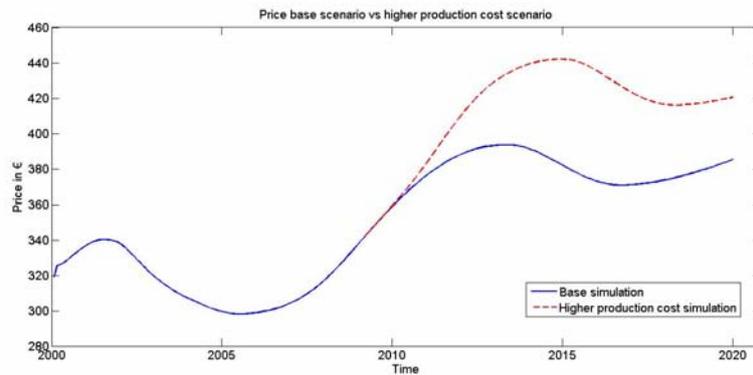


Figure 9: Increased cost scenario

4. Conclusion and further remarks

In this paper a system dynamics model has been developed and was used to simulate the milk price dynamics in different scenarios. The simulation results have shown that the model can fit historical data and because of that should do a good job in forecasting the dynamics of the milk price.

Nevertheless the model suffers several limitations. The following gives an incomplete list of the key items which will be addressed in future research:

Chapter 0. The model is calibrated to yearly data. Due to this relative long time interval between the data points it is difficult to capture the short time dynamic of the milk price. In addition relative few data points are available to calibrate the model. This problem will be addressed in the future by switching to monthly or quarterly data.

Chapter 1. In the model milk demand is modelled at an aggregate level. This means the demand for the major dairy commodities (e.g. cheese and butter) is summarized in one variable. Studies show that the price of dairy commodities influences the milk price. Explicitly modelling the dairy commodity markets will address the relationship between these markets and thus improve the forecasting quality of the model.

Chapter 2. The model assumes that farmers behave fully rational and set future supply based on profitability. The milk crises starting 2007 showed that farmers are not necessarily rational, which means they rather dumped milk into fields then to sell it for a low price. Incorporating insights from the behavioural science discipline might improve the forecasting quality of the model and could give new insights.

5. Appendix: Symbol description Anylogic

Figure 10 gives a short description of the different variable types used in system dynamics. Parameter came from outside the model and need to be estimated. A stock is the value of a resource fixed at a specific point in time (e.g. the level of water in a bathtub). A flow changes stocks (e.g. the flow of water in and out from a bathtub) and auxiliary variables are functions of stocks or flows.

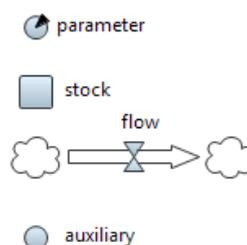


Figure 10: Anylogic symbol description

6. References

- EU Commission (2010), "EU dairy farms 2010", http://ec.europa.eu/agriculture/rica/pdf/dairy_report_2010.pdf, (Accessed 20. Jan. 2012)
- FAPRI (2011), "FAPRI-ISU 2011 WORLD AGRICULTURAL OUTLOOK", <http://www.card.iastate.edu/publications/dbs/pdffiles/11fsr1.pdf>, (Accessed 20. Jan. 2012)
- Mankiw, N.G. and Taylor, M.P. (2008), "Grundzüge der Volkswirtschaftslehre", *Schaeffer Poeschel*, ISBN 978-3-7910-2787-6
- Meadows, D.L. (1970), "Dynamics of Commodity Production Cycles", *MIT Press*, ISBN 0-26-2131412
- Morecroft, J. (2007), "Strategic Modelling and Business Dynamics: A Feedback Systems Approach", *John Wiley & Sons*, ISBN 0-47-0012862
- Osorio, F.A. and Aramburo, S.A. (2009), "A SYSTEM DYNAMICS MODEL FOR THE WORLD COFFEE MARKET", *Proceedings of the 27th International Conference of the System Dynamics Society*
- Sterman, J.D. (2000), "Business dynamics", *Irwin/McGraw-Hill*, ISBN 0-07-238915X
- Xjtechnology Corporation Web Site (2012), "Why Anylogic simulation software", http://www.xjtek.com/anylogic/why_anylogic/, (Accessed 20. Jan. 2012)

Political Marketing Segmentation within English Local Government Elections

R. L. Tidy

School of Management
Plymouth University, Plymouth, United Kingdom
e-mail: rebecca.tidy@plymouth.ac.uk

Abstract

Despite the consistency of academic approaches to marketing segmentation, too few political parties are segmenting their markets effectively. Therefore, Bannon (2004) has developed a model that can be used in order to segment the electorate into different social-groups based upon their behaviour and attitudes towards elections. This paper aims to test the validity of his model by utilising the results of a participant observation and series of qualitative interviews that were carried out during the run-up to the 2011 local elections.

The results suggested that whilst the model can provide an overview of how to segment and target individuals; it was often the case that the political party did not have the time or financial resources to fully utilise it. The respondents under study did attempt to build and maintain relationships with key target groups, such as activists and supporters. However, it was clear that little real targeting took place, relative to their mass campaigning efforts.

Keywords

Political marketing segmentation, Local elections, Case-study,

1. Introduction

Marketing segmentation, targeting and positioning techniques are commonly used within the political arena (Lees-Marshment, 2001; Bannon, 2004; Wind, 2000). There is an abundance of academic literature detailing the various aspects of this process, the perceived advantages, and perceived disadvantages. Despite this, there appear to be very few case-studies, and, similarly, very little literature concerning the implementation of these techniques, particularly within the UK, hence meaning that there is actually very little empirical evidence that details how and even, if political parties in England are using these basic consumer marketing techniques in order to segment the electorate. This means that despite the consistency of academic approaches to marketing segmentation and the theoretical acceptance by political parties of the value of marketing segmentation, it could still be argued that political organisations are not segmenting the electoral market effectively and may not be basing their campaign strategies upon research-based segmentation.

Therefore, this working paper reports upon the beginning of an attempt to fill that gap and assess the extent to which political parties are utilising segmentation, targeting and positioning strategies within the electoral marketplace. It is commonly suggested by (Bannon, 2004) and several others (Wind, 2000; Bradshaw, 2005; Baines et al; 1999) that the constituency organisations of political parties are implementing some of the principles of these segmentation strategies, however, he offers no evidence-based research to support this theory, such as the detailed strategies used by well-known brands, such as Habitat, Jessops, and so on (Baker, 2000).

The aims of this stage of the project are to:

- Assess the extent to which the Liberal Democrat party are utilising segmentation techniques as part of their marketing strategy
- Generate a body of empirical evidence to either refute or support the claims of (Bannon, 2004) and (Wind, 2000)

The results are based upon data generated during a participant observation and series of semi-structured interviews, both carried out between March and May 2011. The case-study was carried out in 3 local government wards within an urban parliamentary constituency. The constituency organisation has asked to remain anonymous. Additionally, a secondary objective was to discuss and explore the potential benefits and limitations of the marketing segmentation, by conducting a brief literature review.

2. Literature Review

2.1. The English Political System and Environment

England, like the rest of the UK, has a majority system; the party elected to government is the party that wins the most seats in a general election, decided by a first-past-the-post electoral system held at least every 5 years. The system is dominated by 2 main parties; the Labour party and the Conservative party. It has functioned as a duopoly for the greater part of its existence and third-parties have generally been marginalised. The main chamber in the legislature is the House of Commons, to which all MPs are elected by their constituencies. The upper chamber; the House of Lords is unelected and acts as a check on government legislation.

The main organisations compete at all elections, employ professional advisors and staff, but also utilise voluntary labour through membership. Membership remains an important part of the party organisation in Britain. As only 2 parties dominate politics, they gain more media attention, though this means that every aspect of their behaviour can be conveyed nationally and therefore, has potential influence on voters. The political system and the comparatively high levels of party unity enable a prime minister elected with a reasonably good majority to pass policies relatively easily through Parliament. The scope for party leaders to utilise political marketing is therefore, very high.

The Conservative party is the oldest established UK party, with a tradition going back to the 17th century. The party have evolved from a party of landowners, the rich and the aristocracy, to court the expanding middle classes, and eventually, the skilled working classes as the franchise expanded. In contrast, the Labour party was formed in 1899 as the political arm of the trade unions. It adopted labourism as the platform from which to represent the working classes in the 1920s and by the 1930s had established itself as the second main party. Similarly, Labour won the broad support base; geographically, socially, and occupationally, including the middle classes, as well as the socially deprived, in order to form a government.

Until the 1980s there were no other significant parties. In 1988 the Liberal democrats were formed out of the remnants of the Social Democratic Party and the Liberal party. They have solidified a support base at the local level and offer some challenge to both main parties. Smaller parties such as the Greens and the British National Party have competed in some elections, increasing their overall public support and winning local council seats, but with no success nationally because of the nature of the electoral system.

2.2. Political Marketing

Political marketing is playing an increasing role within everyday political life; parties from all over the western world develop their manifestoes based upon the results of qualitative and quantitative marketing research. Commercial techniques and strategies, such as, focus-groups and telephone polling, are now commonplace in the political arena (Scammell, 1995). It is frequently acknowledged that these more complex techniques are being used in response to the rise of the more critical, better educated and informed voter who is independent of a particular political party (Gibson, 2003; Jackson, 2009). (Lees-Marshment, 2001) goes even further and argues that voters choose parties as consumers choose products. As a result, several academics have examined the extent to which marketing techniques and approaches are employed; in a similar fashion, this paper details the results of the first stage in a research project that aims to examine the extent to which segmentation, targeting and positioning are used. The emergence of the market-oriented party; a party no longer tied to historic ideology, but more focussed upon developing a credible product or service offering that will satisfy the core electoral market is frequently discussed (Lees-Marshment, 2001).

The market-oriented party turns around traditional ideas about politics and argues that in order to win an election a party needs to identify and understand public priorities, concerns and demands, before designing a product that reflects them (Scammell, 1995). It does not try to change what people think, rather, it tries to deliver what they need and want (Lees-Marshment, 2001). UK politics is often described as having a strong market orientation. Margaret Thatcher, prime minister from 1979-1990 implemented a market-oriented approach; this was not just limited to presentation, but also the design of the political product (Scammell, 1995). Similarly, Labour's leader since 1994 and prime minister from 1997-2007 has directed UK politics to a fuller market-orientation, whereby public demands influenced the rebuilding of the entire party product, including emblems and logos, policies, internal structure and even the party constitution (Lees-Marshment, 2001; Scammell, 1995; Bannon, 2004).

2.3. Segmentation, Targeting, Positioning, Implementation, and Evaluation

A consumer market is not a mass of potential customers all with the same values, desires, aspirations and ability to be a customer for any given organisation, and similarly the electoral market-place is not a mass of potential voters all with the same values, desires, aspirations and ability to be a voter for any given political party (Bannon, 2004). However, similarly to customers, voters' attitudes can be grouped into sub-groups or segments of the whole market or electorate (Lees-Marshment, 2001). Political marketing segmentation can thus be viewed as the sub-dividing of the electoral market into groups with similar attributes (Bannon, 2004). A market segment will be valid if a political party decides it is capable of delivering the specific requirements that match the success criteria for marketing to that segment and that the segment is sufficiently large enough to be attractive (Lees-Marshment, 2001).

The major stages of the process are:

- Marketing Segmentation - identify bases for segmenting the market; develop profiles of resulting segments
- Market Targeting - develop measures of segment attractiveness; select the market segments
- Market Positioning - develop positioning for each target segment; develop marketing plans for each segment
- Implementation - development of marketing plans to implement marketing segmentation; evaluate the benefits derived from the activities and refine the process

Reference: Bannon (2004)

The following section of this paper discusses segmentation in more detail.

2.4. Segmentation

The segmentation of a market requires criteria or variables that can meaningfully differentiate elements in a population (Brown, 1998). There are two approaches for segmenting markets; a priori and clustering. Clustering or post-hoc segmentation, as it is commonly known, involves the design of the segments using relevant criteria induced from research in the market being segmented (Bannon, 2004). In some cases, a mixture of a priori and clustering approaches may improve the criteria selected (Baker, 2000; Green et al, 1998; Myers, 1996; Smith and Hirst, 1991; Wind, 1978). Cluster analysis is potentially a very useful technique; however, it can prove difficult in its application (Everitt, 1974) and therefore, is likely to be beyond the scope of the financial and human resources of many constituency organisations of political parties. It is for this reason that this article is focussed upon a priori segmentation.

The second form of segmentation; a priori segmentation, utilises management knowledge of a market to select criteria or variables that can be used to sub-divide a population (Brown, 1998). The criteria are deduced prior to segmentation. Marketers often do this instinctively. Bannon (2004) offers some examples:

- Older people are more likely to vote, therefore, the segmentation criteria chosen may be age
- Former mining towns predominantly vote Labour, therefore, the criteria chosen could be geographic distribution
- Previous Tory voters are more likely to vote Tory, so the criteria chosen is prior behaviour

Political marketers rarely challenge the application of marketing to politics (Lees-Marshment, 2001). However, some politicians and political scientists have reservations, and are often sceptical or even hostile to it. Despite this hostility, all members of the electorate are not equally likely to vote for a party engaged in mass marketing, and this is why segmentation is commonly used (Lees-Marshment, 2001). There are numerous ways to segment markets (Bannon, 2004). Wind (1978) suggested that the theory of marketing segmentation suggests that there is one best way of segmenting each market. Any attempt to use a single basis of segmentation for every single segmentation exercise and subsequent marketing decisions carried out by an organisation, could, therefore, result in inappropriate outputs and wasted resources. Table 1 details common bases for segmentation and how they can aid marketing decision-making.

<u>Segmentation Base</u>	<u>Characteristics of Interest</u>	
Geographic	Values Behaviour and motivation Historical factors Culture and tradition	Postcode of an individual can reveal characteristics that differentiate that person
Demographic	Age Gender Socio-economic basis Family life-cycle	Often associated with consumer or voter preferences
Geo-demographic	ACORN; a classification of residential neighbourhood Provides details on a mixture of geographic and demographic variables Much greater discrimination than occupation based scheme of social grading	Aids and directs campaign activities
Psychographic and behavioural	Individual actual behaviour and lifestyle behaviour Beliefs Attitudes Activities Interests Opinions	Concentrates on the benefits that services bring, not on a functional level, but in an emotional and lifestyle enhancing level

Table 1: Common Bases for Segmentation (Bannon, 2004)

This paper is concerned almost entirely with segmentation. Therefore, it is not necessary to discuss market targeting, positioning, implementation, and evaluation, to a great extent. The extent to which the party are engaging with these techniques will be analysed in future work.

2.5. Arguments against Political Marketing Segmentation

It is often argued that marketing segmentation can be viewed as anti-marketing in that it is not treating the customer as a unique individual but as a standardised unit within a segment (Resnick et al, 1979). Furthermore, there is a tendency to blindly assume that marketing segmentation is a good activity and therefore, not to challenge the supposed benefits. This is naïve and misleading, and it can potentially be an expensive mistake (Cui and Choudray, 2002). In certain cases, the costs of segmentation can outweigh the benefits. In these circumstances it is inappropriate to segment the market (Resnick et al, 1979).

There are a number of practical issues with the marketing segmentation process. Critics often argue:

- Over emphasis upon segmentation techniques can neglect the electoral market itself and the competitive position of the political party competing within it (Bliss, 1980).
- Cost benefit analysis
- Validity and reliability of criteria for segmentation
- Depth of segment, i.e. at which stage do you stop segmenting? Too much segmentation loses some of the benefit and cost advantages of market segmentation (Resnick et al, 1979). This is referred to as counter-segmentation.
- Stability of segments produced
- Ability of the party to tailor services to any identified segment
- Expertise within the organisation

Reference: Bannon (2004)

Overall, it is commonly agreed amongst political marketers, that there are more benefits than disadvantages to marketing segmentation as it can be implemented with varying levels of refinement and sophistication (Baker, 2000; Bannon, 2004; Lees-Marshment, 2001; Yorke and Meehan, 1986).

3. Methodology

A six-week participant ethnography was carried out. Throughout this research, the term ‘ethnography’ has been used in a liberal way, not worrying much about what does, and does not count as examples of it. It involves the ethnographer, in this case overtly, participating in people’s daily lives for an extended period of time. The ethnographer watches what happens, listens to what is said, asks questions, in order to cast light upon the area of research (Hammersley and Atkinson, 1983).

Ethnography was used because it bears a close resemblance to the routine ways in which people make sense of the world in everyday life (Wax, 1971). It has become a popular approach to social research, along with other types of qualitative work, such as interviews and focus groups (Hammersley and Atkinson, 1993).

This involved living with a retired councillor in the ward from 24th March to 5th May 2011. Throughout this period, much time was spent with the candidates, election agent, volunteers, and campaign officer, in order to assess the extent to which the local branch was engaging with the market segmentation process. In order to prevent the loss of any important details, it was necessary to keep a diary. This diary was updated on a daily basis.

The campaign team began work at 8am and often did not finish until 10pm. The candidates, staff and volunteers worked for 7 days a week during the campaign period. In order to immerse oneself fully in the culture of the local organisation and to gain significant depth of analysis, it was decided that it was necessary to be present for 6 of these days and also, between the hours of 8am and 10pm.

Whilst ethnographers may assume a variety of roles, it was decided that a marginal position would be assumed throughout the period, thereby providing access to participant perspectives whilst minimizing the dangers of over-report (Hammersley and Atkinson, 1993).

Semi-structured interviews were also carried out and analysed in order to gain a deeper understanding of the extent to which the local branch of the party were using market segmentation. They also provided a means of triangulation. The interviews were carried out with three council candidates; one of whom was successfully re-elected, one campaign officer, and three volunteers or activists. All of these people lived and worked within the geographic area under study.

The majority of the seven interviews were carried out throughout the local election campaign period; 11th April to 5th May 2011. However, it was necessary for two of the interviews with candidates to be carried out after the election, as they were busy campaigning beforehand.

4. Results and Discussion

This research aimed to answer the following question:

- To what extent did the Liberal Democrats attempt to use marketing segmentation within the council wards in the Bristol South parliamentary constituency during the 2011 local elections?

This section of the paper discusses the extent to which the anonymous branch of the Liberal Democrats featured in this case-study use the four aforementioned types of market segmentation throughout the campaign period leading up to the 2011 English local elections.

4.1. Geographic Segmentation

Throughout the participant observation and interviews it became clear that the local party attempted to use geographic market segmentation. Prior to moving to the area and beginning the research, the campaign manager commented that, "We're always segmenting the local area. It makes life so much easier." Later on, when interviewed, he explained that geography can indicate various factors, including the density of the housing stock in an area. Interestingly, the density of the housing stock was not used to decide voter intention or how likely a person was to turnout at the polling station, it was used to decide which volunteer(s) would be asked to deliver literature in a certain area:

"Lots of our volunteers are older, retired people and they would not be able to walk long-distances between houses. They often struggle to walk up lots of steps. Fortunately, we can look at Google Maps or use our own knowledge to divide a ward up into several geographic areas. This means we can give the bungalows on hilly streets to younger, more able-bodied activists, and avoid scaring the older people away, as their help is still vital to our campaign!"

Various council candidates also mentioned that geography can often provide a rough indicator of historical voting trends. An incumbent candidate who said that she was running in a relatively "safe" ward explained that each electoral district, i.e. ward, was divided into 4 different areas, referred at as, "Areas A, B, C, and, D." There is one polling station in each area. In previous years, she had attended the count with her colleagues and learnt which of the four areas she received the highest share of the vote and which she received the lowest. Whilst attending the counting of the postal votes, which occurred approximately two weeks before polling day, she commented, "We always attend the counting of the postal votes because voting trends in each area do change over the years. We expect them to change a lot this year; mainly, because the national party are in an unpopular coalition."

Following the count, the local team met and decided that they would focus more heavily upon one geographical district, as after watching the counting of the postal votes, the party appeared to be performing particularly badly there. Prior to this, little geographical segmentation with a view to targeting voters had been carried out. The only exception to this was when the local team decided not to target areas with very low levels of turnout. When asked why they chose not to canvass or distribute literature in areas of low turnout in the hope of encouraging them to vote, all three candidates agreed that it would be very difficult to persuade someone that does not generally vote to turnout on polling day or to vote by post. One candidate said that, "In all honesty, it's more effective to spend our limited time communicating with people that intend to vote, and encouraging them to turnout on a rainy polling day. Experience tells me that if they didn't vote in the general election then they definitely won't vote in the locals and we haven't the time..." This suggests that geographic segmentation can sometimes be indicative of behaviour, and this will be discussed later in the paper.

4.2. Demographic Segmentation

One Liberal Democrat candidate said, "I have absolutely no idea what you mean by demographics," when she was asked whether she used demographic segmentation throughout her election campaign. Similarly, the word, "demographic" was rarely used throughout the 8 weeks prior to the election. However, it would be incorrect to assume that the local organisation do not engage in demographic market segmentation. Whether consciously or sub-consciously, they do segment the electorate according to age, and socio-economic bases, for example, by judging the likelihood of a person to vote for the party, simply by looking at cues such as their occupation or appearance. They simply do not apply the term, "demographic segmentation" and it is not carried out formally.

Furthermore, over the 8-week period, it did become apparent that the local organisation used gender, stage of life-cycle, and stage of family life-cycle, to segment the electoral marketplace. This was discussed with the campaign manager in an interview and he said that he felt that whilst demographic segmentation could potentially help the local party to target their messages, it was very difficult to get hold of demographic data relating to individual households. He explained that he wanted to target parents in the ward, as they had spent £30 000 on play-equipment in various parks,

however, instead, he simply added this information to a paper newsletter and distributed it to each house in the electoral district, which he felt was far less efficient as it was not targeted.

Throughout the eight weeks prior to the election, the local branch relied entirely upon putting literature through letterboxes and knocking on doors to talk to potential voters. Other mediums, such as local newspaper advertisements, the use of social-networking sites, and so on, were not really considered. Local volunteers explained that they felt they were either too costly or not effective enough. It would be necessary to carry out further research to see if this pattern is common to the entirety of England and also, other political parties. This research will be carried out in due course.

A candidate in the area explained that, “It is the job of Cowley Street [Liberal Democrat HQ] to carry out demographic research and segment people according to their interests. We don’t engage in that kind of thing. We haven’t the resources... No time or money.” This fits well with previous work by Denver et al (2003) that details the campaigns of the 3 main parties throughout local and national elections in the 1992-2001; they also found that the Head Offices of the various political parties carried out demographic research and targeted various groups of people in target-wards or constituencies, either with direct-mail, billboards, or visits from well-known politicians. Similarly, in past elections party-leaders have visited mother’s groups, or community groups, in order to promote certain policies, and gain media coverage in the local newspapers in various areas (Clark, 2003).

Despite this, the party HQ has announced the introduction of a new database that allows activists and candidates to record demographic details of households, in addition to just the voting intention of occupants. This was announced after the local elections, at a conference intended to provide a debrief and evaluation of candidate performance. The party employee introducing the new system explained that they wanted local branches to improve the way in which they segment, target, and subsequently, position the party in the minds’ of the electorate. He said that the purpose of this was to build up a “relationship” with voters by targeting voters with information about the party’s achievements that have benefitted them, for example, targeting those on lower incomes after the party increased the threshold at which income tax becomes payable, or targeting local parents with leaflets about the improvements to the local park. It was explained that this information could be entered into the database via a Smartphone after a volunteer had knocked on a door, or it could be entered via computer, when the person returns to the office or home. If used properly, this will allow additional types of segmentation to take place.

4.3. Geo-Demographic Segmentation

Similarly, the participant observation showed that the constituency organisation do not use geo-demographic segmentation. Again, they say that a lack of resources prevents this. Interestingly, however, candidates and volunteers have developed their own means of combining geography and demographic information, in order to decide which areas to target. For example, a retired councillor explained that they campaign in most of the ward, then she commented, “We never campaign in [place]; we call it Little Beirut because nobody has a job in those streets, the area is awfully frightening, and the majority of people living there do not vote.” Whilst this is a very sweeping generalisation, it does suggest that candidates improvise their own means of segmentation, often in a rather ad-hoc manner.

One party employee explained that each constituency organisation was offered the opportunity to subscribe to a service that provides regular ACORN reports, so that candidates and their colleagues can map the people in the local area, and gain a deeper understanding of the electorate. He said that the Party HQ believe that ACORN is a very important and useful tool in political marketing, but that many constituency organisations struggle to find people with the skills to use the service offered. Therefore, it could be suggested that the local organisation in this case study could segment more effectively, if they had the perceived ‘expertise’ and time required to make use of these geo-demographic segmentation methods, such as ACORN. Candidates would subsequently have the opportunity to target parts of the ward, potentially by mail, according to their broad social, economic, age, and household characteristics.

4.4. Psychographic and Behavioural Segmentation

Psychographic and behavioural segmentation are both useful because they focus upon an individual’s actual behaviour and lifestyle, rather than trying to infer that certain characteristics will lead to a particular behaviour (Bannon, 2004). Furthermore, political segmentation appears to suit psychographic segmentation because generally the electorate does have opinions, beliefs and attitudes towards political parties.

(Bannon, 2004) suggested that there are three divisions of social character; traditional directed behaviour, i.e., a voter always votes this way; conformity with individual or family groups, i.e. socialisation into supporting a party, and, individualist behaviour, i.e. instrumental consumer buying behaviour of choice of party.

Constituency organisations are encouraged to subscribe to a service named EARS, a database that allows volunteers to record the voting intention of local residents. Groups of volunteers and candidates use a map and knock on doors in the ward that they are targeting. They then record the voting intention of respondents. There are several reasons why constituency organisations engage in this type of segmentation. A local activist said:

“We knock-up [knock on doors and remind people to vote and offer to take them to the polling station if they have no transport] from dawn to dusk on polling day. We don’t have time to knock on everyone’s door and we don’t want to remind supporters of the Labour party or the Conservatives to vote either. Therefore, we print stuff out of the database and knock on the doors of our supporters to ensure that they turn out.”

Throughout the eight weeks, it was clear that segmenting the electorate in this way fulfilled several purposes; firstly, it allowed for the identification of non-voters. This meant that the local organisation did not waste time or money sending leaflets or related literature to them. A visiting MP said, “If they say that they didn’t vote in the general election, then we assume that they will not vote in the local elections.” In traditional product- or service-based marketing, this would be referred to as “usage rate”. Secondly, it identifies people that are considering voting for the party or similarly, are unsure who to vote for, and this means that they can be targeted with selected mailings designed to encourage the elector to vote for the Liberal Democrats. For example, several leaflets and a personalised letter were delivered to all respondents that had said they were not yet sure who they planned to vote for, or that they were considering voting for the Liberal Democrats. These leaflets included information concerning the incumbent’s achievements in the local area, or what the challenger Liberal Democrat candidate would do, if they were elected, and also, details concerning the policies that the national party had implemented since being in coalition with the Conservative party.

Respondents who said that they planned to vote for a different political party received no more literature or communication. The campaign manager explained;

“There is one section of the ward where literally an entire housing estate votes Labour. We never deliver anything to them in the run-up to the election. We don’t want to remind them that there is an election; we just hope that they forget.”

Voter-loyalty also plays a role in segmenting the electorate. For example, if a person says that they are considering voting for another party, such as, the Conservative party, then they will subsequently receive mailings attacking that party and criticising what they have, or have not, done for the local area.

Naturally, one of the most clear examples of behavioural segmentation, and subsequent targeting, and positioning is the targeting of specific wards. This is usually achieved by identifying marginal wards. This is encouraged by the aforementioned first past the post voting system where parties aim to maximise the number of seats won, not votes won. Resources are subsequently allocated towards these targeted marginal wards or seats.

Despite being regarded within academic literature as relatively complex (Bannon, 2004; Lees-Marshment, 2001), a Liberal Democrat employee explained that psychographic and behavioural segmentation are commonly used by constituency organisations across the country. He said that it was generally thought that it provided a more meaningful method of segmenting the electorate; he explained, “This is because it focuses upon a voter or non-voter’s actual behaviour, rather than trying to infer that certain characteristics will lead to a particular behaviour, for example, people living in old industrial working-class towns predominantly voting Labour.”

5. Conclusion

This paper comes to a rather unsurprising conclusion; in that the Liberal Democrats within this particular geographical area are implementing some of the most basic principles of market segmentation. These include segmenting according to geographical area, in order to allocate more hilly and difficult streets to fitter volunteers, hence increasing campaign intensity and efficiency, using geographical segmentation as an indicator of voter intention, by viewing the results of the counting of postal votes in order to increase campaign efficiency. Despite this, not all of the candidates attended the postal vote count for their seat, and two of them simply continued to deliver a generic piece of literature to every part of the ward. This suggests that there is significant scope for improving segmentation and subsequent targeting and positioning, within electoral districts, in order to maximise vote-share.

Similarly, there is significant scope for incorporating demographic segmentation into the local planning of campaign strategy. Candidates and volunteers often “confessed” to judging voters by their occupation or appearance, however, none of them used these criteria in order to segment the electorate. Additionally, when canvassing; volunteers chose to mention certain achievements and policies to voters, dependent upon whether they had young children and used the local park, and so on. If this information was to be recorded and voters subsequently targeted with relevant literature, it could aid a relationship marketing strategy and ultimately, increase vote-share in the electoral district being targeted. A

candidate summed up the general view towards demographic segmentation; “I have absolutely no idea what you mean by demographics.”

As previously discussed, geo-demographic segmentation is not used within the area. Despite acknowledging the potential of this method, the constituency organisation felt that they did not have the expertise and time required to do this effectively. Several forms of segmentation are frequently used by various members of the constituency organisation; these include psychographic and behavioural segmentation, both of which are used more extensively than the aforementioned bases. This was believed to be particularly efficient as it allows for the identification of non-voters, switch-voters, and so on, thus allowing time and resources to be utilised more effectively.

Of course, these results cannot be generalised to all constituency organisations operating across England, and cannot even be generalised to all political parties within the country. Further semi-structured interviews will be carried out in order to enable a significant depth of analysis. Similarly, another participant observation will uncover the extent to which political parties within England use targeting and positioning, as part of their marketing and campaign strategies. Additionally, the role that Web 2.0 and other forms of new media can play in this process will be examined.

In order to address the limitations of this study, quantitative research is planned. An online survey of local election candidates will be carried out in the run-up to the 2012 English local elections. One in six candidates will be sent a postal invite to participate in an online survey. It is anticipated that the results of this survey will be used alongside the qualitative research in order to provide detailed analysis within this area.

6. References

- Bannon, D. P. (2004), “Marketing Segmentation and Political Marketing”, *Proceedings of the PSA Annual Conference*; University of Lincoln, Lincoln.
- Baines, P. Lewis, B. and Yorke, D. A. (1999), “Marketing Planning for UK Political Parties: Co-ordinated Local Campaigning”, *Proceedings of the Marketing Academy Conference*, Stirling University, Stirling, July 1999.
- Baker, M. (2000), *Marketing Strategy and Management* 3rd Edition Macmillan Business pp. 461-462.
- Bliss, M. (1980), “Market segmentation and environmental analysis”, unpublished MSc thesis, University of Strathclyde.
- Bradshaw, J. (2005) “Who will vote for and why: designing strategy and theme” *In Campaigns and Elections: American Style*. (Eds) Thurber, J. A. and Nelson, C. Westview Press, Boulder, USA.
- Brown (1998) *Postmodern Marketing 2: Telling Tales* Thomson Business Press, London.
- Clark, A. (2007) “Community Politics? Grassroots Liberal Democrats and the 2003 Scottish Parliament Elections”, *British Journal of Politics and International Relations*. 9 (1), p696-712.
- Cui, G. and Choudray, P. (2002), “Marketplace Diversity and Cost Effective Marketing Strategies”, *Journal of Consumer Marketing*, Vol. 19, No. 1. pp.54-73
- Denver, D. Hands, G., Fisher, J. MacAllister, I. (2003). “Constituency Campaigning in Britain 1992-2001: Centralisation and Modernisation”, *Party Politics*. 9 (5), p541-559.
- Everitt, B. (1974), *Cluster Analysis*, Heinemann Educational Books, London.
- Gibson, R. K. Margolis, M. Resnick, D., Ward, S. J. (2003). “Election Campaigning on the WWW in the USA and UK: A Comparative Analysis”, *Party Politics*. 9 (1), p47-75.
- Green, P. Tull, D. and Albaum, G. (1988), *Research for Marketing Decision*, 5th Edition, Prentice-Hall, Eaglewood Cliffs, NJ
- Hammersley, M. and Atkinson, P. (1983), *Ethnography*, London: Routledge.
- Jackson, N. and Lilleker, D. (2009), “MPs and E-representation: Me, MySpace and I.”, *Party Politics*. 4 (2), 236-264.
- Lees-Marshment, (2001), “Political Marketing and British Political Parties; the Party’s Just Begun”, Manchester University Press; Manchester.
- Myers, J. H. (1996), “Segmentation and Positioning for Strategic Market Decisions, American Marketing Association”, Chicago, IL
- Resnick, A. Turney, P., Mason, J. (1979), “Marketers Turn to Counter Segmentation”, *Harvard Business Review*, Sept-Oct, pp100-106
- Smith, G. & Hirst, A. (2001), “A New Approach for a New Era of Political Marketing”, *European Journal of Marketing*, Vol. 35, No. 9/10 pp. 1058-1078

Political Marketing Segmentation within English Local Government Elections

Scammell, (1995), "Designer Politics: How Elections are Won", Macmillan Basingstoke, Hampshire and New York, N.Y.

Wax, R. (1971), *Doing Fieldwork: Warnings and Advice*, Chicago, University of Chicago Press.

Wind, Y. and Cardozo, R. (1974), "Industrial Market Segmentation", *Journal of Industrial Marketing Management*, Vol. 6, No. 2, pp. 17-25.

Wind, Y. (2000) "Marketing Segmentation" in *Marketing Theory*, edited by Michael Baker, pp.181-205, Thomas Learning, London.

Yorke, D. and Meehan, S. (1986), "ACORN in the Political Marketplace", *European Journal of Marketing*, Vol. 20, No. 8, pp. 63-76.

Chapter 4

Electronic Engineering

Overall Transparency in Distribution Planning using Software-based Solutions and Low-cost Inertial Navigation Systems

M. Haid¹, M. Kamil¹, T. Chobtrong¹, E. Günes¹, M. Münter¹, H. Tutsch²

¹ Competence Center for Applied Sensor Systems (ccass)
Hochschule Darmstadt – University of Applied Sciences, Darmstadt, Germany
e-mail: markus.haid@h-da.de

² Flexis AG, Research and Development, Stuttgart, Germany
e-mail: hansjoerg.tutsch@flexis.de

Abstract

Vehicle ID and location information in distribution processes are the key information for connecting planning and logistics. Through this coupling the logistic process is provided with object-related data which answer the question about which vehicles are currently where positioned on production site. The integrated use of location and job information (steps to execute for the vehicle finishing, e.g. refuelling, washing, packing, etc.) in one system allows closing the information gap between planning and implementation of logistics which is currently still available in involved companies. Thus, it is for example possible to react quickly and accurately on any unforeseen changes in the process - be it through road works on sections of transport routes or on urgent requests. The cooperating partners are developing an integrated tracking and planning solution for the localization of vehicles on factory sites. Particularly, position information allow identifying failure and disturbances in the process which can be employed by a planning software to initiate quick and efficient reactions. Here, a low-cost navigation module can offer the possibility of a flexible and sustainable adaption of planning, scheduling and control on the needs of turbulent processes.

Keywords

Efficient Distribution, Automotive Logistics, Object Tracking, Inertial Navigation, Navigation in Difficult Environments, Road Sign Recognition

1. Introduction

In distribution logistics of automotive plants, the transport of finished vehicles to the various loading stations (e.g., train, ship, truck) or to the customer is of great importance for the manufacturer. The transport of products has to be processed quickly and efficiently in order to avoid negative feedbacks on production and also to guarantee the fastest possible delivery to the customer. In the current situation it is not possible for the planners to get a feedback on the implemented work in the distribution process. Next to that it is unknown, if any sudden disturbances have occurred or which need is given for everyone of the available resources. The key information to fill this information gap would be accurate ID-related location information provided in real-time for every vehicle on the ground. In combination with data on tasks to be accomplished and resources available, the tracking of the vehicles can help to add control possibilities to the planning and also gives a powerful tool for the verification of the planer's success. For example, it can be determined at anytime, if the object is transported correctly to the defined target and additionally if this was accomplished in the designated time. Loss and interchanges of vehicles can be indicated immediately, and so, quick and precise reactions can be initiated. Using location and job information in one integrated environment allows for the first time creating a synchronized information base for planning and implementation of logistics, which in the sequel can be used for optimizing the delivery procedure and for quick and accurate reactions on any unforeseen changes. The results are less loss of time, more efficient and lock-free use of resources (e.g. personnel, facilities, transport routes, etc.) and consequently higher productivity.

The tracking and localization solution which is given by the infrastructure developed in this project can be used across different industry branches and applications. For example, it can help to determine the position of forklift trucks in warehouses, security personnel on company grounds or even parts, assemblies and products in production halls.

2. Integrated Planning and Navigation System

The cooperation partners are developing an integrated planning and navigation system with a software part that offers a site-wide environment for the management of the distribution process. The solution combines job information with up-to-date location data of all vehicles on the ground and, thus, allows the user to realize optimal planning of currently available resources. Next to that the solution is also aiming to offer a possibility to plan the future employment of the

facilities to avoid any congestions or bottlenecks, before those happen. The tracking of the vehicles is realized with a sensor platform that is developed by the Competence Center for Applied Sensor Systems (CCASS) at the University of Applied Sciences in Darmstadt. The solution is mainly based on a low-cost inertial navigation technology made of micro-electro-mechanical sensors which are highly available on the market. Because of their small packaging they are ideal for the creation of miniaturized measurement modules. Inertial navigation systems (INS) are reference-less working devices which do not need neither external signals nor expensive infrastructure to operate. INS are highly flexible to be deployed in various environments, such as faraday cages, hot, moist or aggressive atmosphere, shaded sectors in production halls or in large areas and for long movement processes. Moreover, they do not interfere with other units in the process, so that many modules can work next to each other properly. Low-cost INS allow reducing operation cost and accurate tracking for a broad field of applications in which other systems come to their limits. With the presented system it is possible to build up the bridge between production and distribution processes, making expensive and time-consuming installations of two separate tracking systems unnecessary.

Figure 1 shows the general concept for the present research project. The way to track vehicles on a production site leads through the provision of a capable sensor system that can operate in variable environment conditions over long periods of time. It is also known, that the cars sometimes stay parked for more than fourteen days, which forces the development of highly energy-saving circuits and devices.

The information collected by the sensor system are intended to be calculated to position and orientation data on the platform itself. From there the datasets for every vehicle will be transmitted to locally available antennas via wireless connection (Wi-Fi). The latter transfer the data to a central computer which is equipped with a site-wide control and planning software. From there the planners will have the possibility to organize the contemporary given tasks, to analyze the performance of the process and also to study “what-if”-scenarios by varying the process parameters.

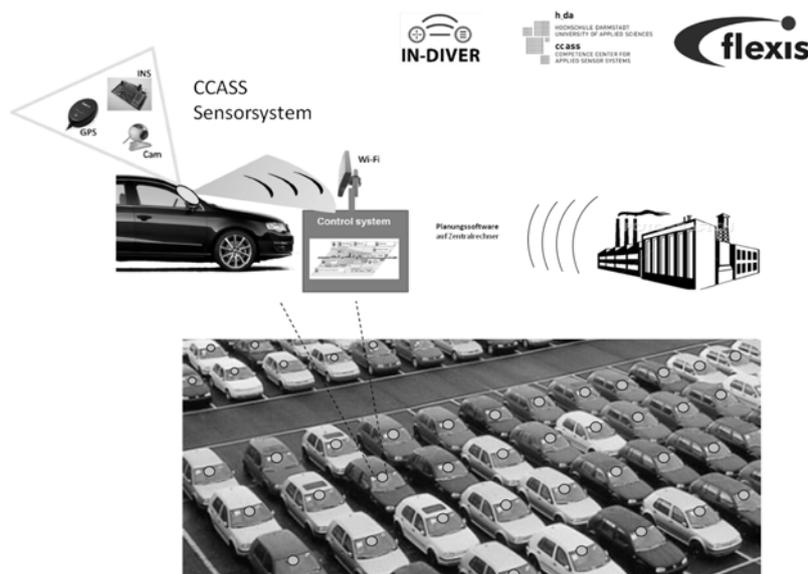


Figure 1: Process Model for Sensor-based Vehicle Tracking and Distribution Planning (CCASS, 2012)

The task is extremely complex and at the same time it treats a highly relevant problem especially in car manufacturing industry. The regulation of processes can only be based on accurate position data of the vehicles and on the fusion of IT tools with a flexible tracking solution in one system.

2.1. Inertial Navigation

The principle of inertial navigation is based on the measurement of object movements with the help of mass inertia under acceleration. For this, an orthogonal constellation of both three acceleration and three angular rate sensors is needed. This assembly allows the determination of all accelerations and angular moments applied on any object moving in space, without requiring signals from the surrounding environment. The acceleration sensors are used to capture translational movements in the three spatial directions. The angular rate sensors (gyroscopes) capture the rotational speeds of the three spatial axes (cf. figure 2).

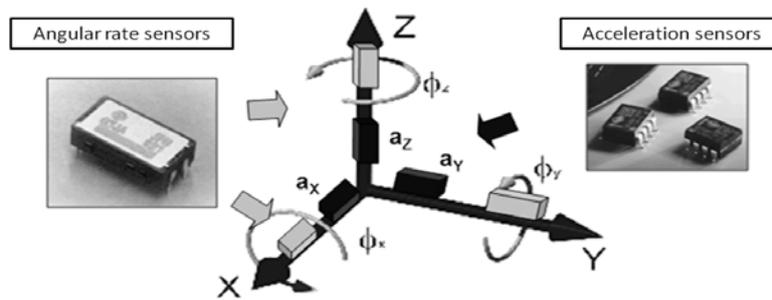


Figure 2: Working Principle of an Inertial Navigation System (Haid, 2004)

Using low-cost sensors for this issue allows on one side a substantial reduction of system cost, but, on the other side, it is also connected with loss of accuracy, accumulating over time. The reason for that is mainly given by a random bias error in the signals of those sensors. With an integration of translational acceleration or angular rate signals this error gets amplified resulting in a time-dependent growth of inaccuracy (cf. figure 4). The scenario of an inertial object tracking on the base of just integrated sensor signals is therefore not possible at the current moment and requires higher techniques to reduce the unwanted side-effect.

In practice, INS are often coupled to other localization systems, for example with a Global Positioning System (GPS) receiver device that periodically provides absolute position data while the INS is used to interpolate the intercostal values. Furthermore, advanced signal processing algorithms can help to reduce the error of position and orientation to achieve a better time performance. For that it is usual to make use of estimation filters like the Kalman filter and to eliminate additional errors evoked by parasitic effects like G-acceleration and the Coriolis effect. In addition to that the use of extra sensors on the inertial platform can provide helpful information, such as the measurement of the earth's magnetic field.

In the presented project, planning and process information also can be used for supplying reference values to the INS. In case that it is scheduled for a vehicle to arrive at a certain time at a specific location (e.g. a gas station, car washing facility, etc.) with known coordinates and that this vehicle is going to stay there for a certain time to pass through an intended finishing step, coordinates calculated by the INS, which will point at a location in the vicinity of this facility can be replaced by the known and exact coordinates of the location and, thus, the INS will automatically be recalibrated. Moreover, it is intended to support the measurement with the information of redundant tracking system to maximize overall accuracy and long-time stability. Table 1 shows a comparison of possible sources of reference positions which were evaluated to support the INS.

Reference System	Accuracy	Rating for vehicle tracking
In-car GPS receiver device	Medium, 12m	Tiny and low-cost, but minor shading problems and only medium accuracy
In-car RFID reader device	Medium, 10m	Tiny and in-car, but no availability of low cost readers with range > 10cm
Laser tracking system on vehicle	High, sub-mm	High accuracy, but large construction and shading problems
Fixed camera based system	High, sub-mm	Shading problems, fixed installation, not in the vehicle
In-car camera based system	High, sub-mm	Ideal for scene/sign recognition, but shading problems
Low-cost ultrasound tracking	Medium, 1cm	Gate-based solution, but no ID-information and not in the vehicle
GSM-based tracking	Low, 50m-150m	Tiny and low-cost, but low accuracy and net registration is needed
Wi-Fi-based tracking	Medium, 1-3m	Tiny and low-cost, but known infrastructure is needed

Table 1: Evaluated Reference Systems in Comparison

As given in the table above, experiments at the CCASS have shown, that the use of both a GPS receiver device and a mobile camera module offers the possibility of integrating the measurement components into the tracked vehicle without the need of an external infrastructure. As a result of the tiny hardware it is also possible to offer a miniaturized platform which makes extra installations unnecessary. It is also shown, that an overall accuracy of at least 12 m can be ensured via sensor fusion. A utilization of a mobile RFID-reader device is ruled out because of the poor availability of higher range device in the low-cost sector.

2.2. General System Concept

On hardware level the INS principle is integrated in a miniaturized measurement platform which is combined with an in-car camera module and a GPS receiver. The resulting sensor system does not exceed a packaging size of 17x10x5cm at the moment and can easily be attached to the windscreen of each car from the inside, where it is protected against moisture, dust, theft and mechanical damage. In a second step it is intended to achieve a higher miniaturisation with the help of a special circuit board design. A wireless networking module is intended to transmit the acquired sensor data from the inside of the vehicle to the Wi-Fi antennas in real-time. The information is then used to fill out a data base provided by the software system developed by flexis AG (cf. figure 1).

2.3. Road Sign Recognition based on Image Processing

Recognizing road signs that are installed on the distribution area can be a source of reference position data for the INS, if the coordinates of those signs are known to the system. This way, it is possible to mark important buildings, crossings and parking spaces with such road signs to ensure the localization of the vehicle before entering them. According to [3] road sign detection is part of current research activities in the automotive industry and there is usually employed to support the adaptive cruise control (ACC) and other driver assisting systems in modern cars. The main difficulty in realizing such a concept is given by the dependency of the contrasts in the acquired images of the lighting situation on the sign and the angle of view. The first, especially, is the decisive parameter which allows the separation of the foreground from background of the road sign, the step that is needed to extract the text printed on it. Another difficulty is given by the possible shading of the signs by snow, dirt, dust or other vehicles which also can be the reason for a failure of the measurement. Nevertheless, multiple experiments have shown promising results in reading different characters and numbers from the video signal which was captured from the inside of the car. According to Figure 3 this was done by attaching the measurement module with the camera to the windscreen to gain a forward view to the signs on the road side. By the use of limitation parameters for the OCR algorithm it could be achieved, that the reading is only executed at a specific distance to the sign, so that this information is provided to enhance the tracking accuracy.

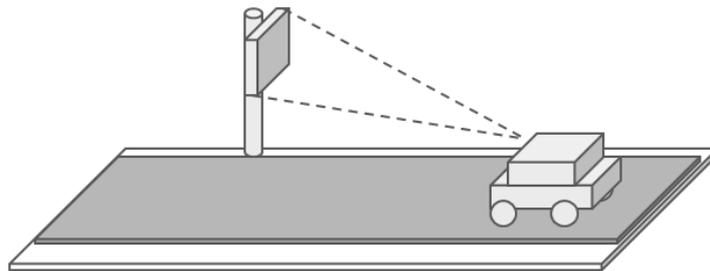


Figure 3: Working Principle of Road Sign Recognition based on Image Processing (CCASS, 2012)

2.4. Analysis

The CCASS has used the campus of the University as a testing field for the evaluation of the presented sensors to analyse each one's measurement characteristics in the given application and its accuracy. Therefore a test vehicle was equipped with a black box that contained an inertial measurement platform, a GPS receiver device and a miniature camera module. Figure 4 shows exemplarily a comparison between the INS and the GPS signals on one of various test routes the experiments covered. It can be seen, that both sensors are not accurate enough to keep the line of the road. The GPS track leaves the road borders several times while the processed INS signal, as expected, shows a growing error over time resulting in a deviation of more than 130 meters after 101 seconds of driving, despite the already integrated compensation of static acceleration through gravity.

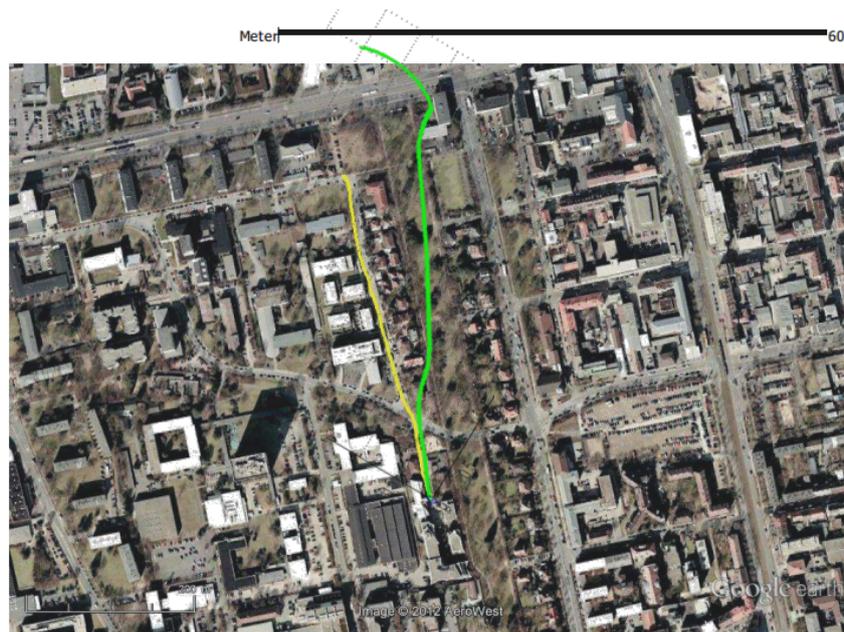


Figure 4: Results of GPS (yellow track, left) and INS (green track, right) Measurements in Comparison (CCASS, 2012)

The road sign recognition algorithm shows encouraging results. In defiance of weak day light and slight rain the characters can be read by software. This was achieved by the use of an optical character recognition (OCR) algorithm which could be optimized by finding optimal values for the threshold between the black foreground of the printed characters and the white background of it (cf. figure 5). Moreover, it was needed to supply the script with fitting template for the printed characters to allow the integrated pattern matching step to provide the desired results. It was noticed, that the recognition became better with every new character template that has been put in.



Figure 5: Sample Result of Road Sign Recognition from In-Car-Video with a CCASS-Software (CCASS, 2012)

2.5. Sensor Fusion

As the experiments have shown, none of the used sensors is good enough to achieve the intended project objectives. The road sign reading via video camera is strongly dependent from variable lighting and weather conditions and also may fail if the free view line to the sign is not given. The GPS readings do not offer a guaranteed accuracy of more than 12 meters of position deviation and also suffer from additional errors in cause of signal reflection. Next to that it is not possible to acquire GPS signals in roofed areas like car washes or filling stations. INS, as described above, have the problem of sensor drifting and as a result the loss of accuracy over time. Additionally INS are mechanical sensors, and hence, can be influenced by shocks and vibration. For the realisation of an appropriate solution for the given task, the future work will concentrate on the implantation of both algorithms for sensor fusion and also for the minimisation of the listed sources of error and disturbance (c.f. figure 6). Mainly it is intended to develop a powerful estimation filter for the inertial measurement and to find a way to adjust the OCR parameters automatically with changing light.

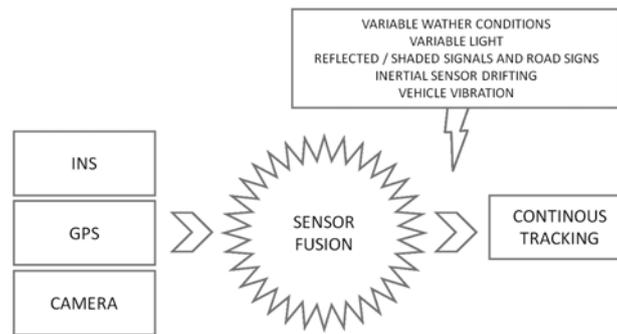


Figure 6: Sensor Fusion Model and Disturbing Factors (CCASS, 2012)

3. Summary

In machine and plant engineering, production logistics and corporate transportation and delivery logistics, tracking technologies open great possibilities to realize optimization and cost saving potentials which, as a result, offer shorter return-of-investment periods for the manufacturers, having a positive effect on investment decisions. For the achievement of these goals, a convenient IT solution has to be combined with a flexible, accurate and performant sensor technology. The usage of inertial sensors for the navigation of vehicles in the distribution process is coupled with many challenges such as sensor drifting and the needing for a suitable data processing concept consisting of estimation algorithms and also of a reference system to minimize the consequences of that issue. The project partners have already developed important parts of the sought system and with the industry-specific software know-how of the flexis AG and the experience of the CCASS in the field of inertial navigation are intending to provide a new innovative and integrated solution for optimized and efficient logistics in the manufacturing industry.

4. References

- Haid, M. (2004), "Improvement of the reference less inertial object-tracking for low-cost indoor-navigation by kalman filtering (Verbesserung der referenzlosen inertialen Objektverfolgung zur low-cost Indoor-Navigation durch Anwendung der Kalman-Filterung)", *Ph.D. dissertation*, Universitaet Siegen.
- Haid, M., Breitenbach, Jan (2004), "Low cost inertial object tracking as a result of Kalman filter", In: *Applied Mathematics and Computation*, Volume 153, Issue 2, ISSN 0096-3003, ELSVIER, Science direct.
- Schreiber, M. (2012), "Conceptual design and evaluation of an approach to a maneuver based vehicle guidance in the context of use highway driving (Konzeptionierung und Evaluierung eines Ansatzes zu einer manöverbasierten Fahrzeugführung im Nutzungskontext Autobahnfahrten)", *Ph.D. dissertation*, Technische Universitaet Darmstadt

Bolt-Identification using an IMU with Bayesian Decision Theory

T. Chobtrong, M. Haid, E. Günes, M. Kamil, M. Münter

Competence Center for Applied Sensor Systems (ccass)
Hochschule Darmstadt – University of Applied Sciences, Darmstadt, Germany
e-mail: thitipun.chobtrong@h-da.de

Abstract

This paper presents a simplified bolt-identification algorithm using a low-cost inertial measurement unit (IMU) signal based on Bayesian decision theory. This algorithm is part of tool-tip tracking algorithms, which will be implemented in an intelligent screwdriver for automotive manufacturing. Because of sensor drift, the accuracy of the position estimation using IMU signals is not sufficient to achieve the target. Based on Bayesian decision theory and the error module with Kalman filter, the bolt-identification module reduces the effect of sensor drift, which is generated by the stochastic noise from the low-cost IMU. The performance of the algorithm was tested with Monte Carlo simulation technique, and also experimentally tested in a lab environment. This bolt-identification module is successfully able to detect a target bolt within its requirement.

Keywords

Bayesian decision, IMU, Pattern recognition, Indoor-navigation

1. Introduction

Quality control is one of the major focuses in automotive industries. Most processes in automotive manufacturing have been carefully observed and tested to ensure the quality of the end products. However, many automotive parts still have been assembled without any quality control process. For instance, all bolts of automotive parts should be fastened in the right position, and also in the right order. This manufacturing process remarkably depends on operators. Therefore, some mistakes might occur by human error, such as missing bolts or fastening a bolt in a wrong position. To prevent these problems and improve the quality of automotive assembly, the system which is able to track the location of the tool-tip location, while fastening the bolts, is required.

The bolt-identification algorithm has been constructed for supporting the project "INSCHRAV", which aims to develop a tool-tip tracking system using a low-cost inertial measurement unit (IMU). One of the requirements of this application is to detect a bolt position with maximum tolerance 10 cm in each direction. This identification algorithm is integrated in an intelligent screwdriver to use it in an automotive production line, such as engine assembly and air-bag assembly. With WIFI communication, the information from the intelligent screwdriver will be continuously sent to the monitoring system for checking the position and the order of fastened bolts.

A low-cost IMU is micro-electromechanical systems (MEMS) composed with three accelerometers and three gyroscopes that measure the linear acceleration and angular velocity in 3-axis of the body-frame (Titterto and Weston, 2004). Theoretically, the attitude of an object attached with an IMU is able to be numerically estimated by integration of angular velocity signals, and the position of the object is estimated by double integration of acceleration signals. However, the major problem of a low-cost IMU is stochastic sensor drift. The error from attitude and position estimation using IMU increases over time dramatically, because of the accumulated error from the integration (Haid, 2003).

Without global positioning systems (GPS), there is a few of successful algorithms to estimate the position of an object using IMU. In (Barshan and Durrant, 1995), the authors developed the error model to generate the optimal error from the result of their experiments. Similarly, Hegrenses *et al.* generated the error module based on KF to compensate the error from position estimation using IMU (Hegrenses *et al.*, 2007). With signals from redundant accelerometers and gyroscopes, Haid successfully applied error module based on KF for one-dimensional position estimation in (Haid, 2004). Interestingly, these techniques are practical when the dynamic model is lack in observation or using IMU without other reference sensors.

Won *et al.* applied a position sensor to improve the performance of position estimation based on KF in (Won *et al.*, 2009). With help of Fuzzy logic algorithms, the position of a fastening tool is successfully tracked with their system.

However, using a position sensor is inconvenient in the automotive production line, because the workspace of this system is limited by the wire length of the position sensor.

Unlike other studies, Vissiere *et al.* developed a KF technique using the magnetic disturbance in observation model to compensate the error of the velocity estimation (Vissiere *et al.*, 2007). Importantly, the performance of this method is dependent on the model of magnetic field disturbances. Based on Maxwell's equation, the mathematic model of this technique is remarkably complicated and difficult to implement in real applications.

To identify the bolt-position that a tool-tip arrives, the classification algorithm is required. Bayesian classifiers are simple to be implemented and good performance (Langley *et al.*, 1992). Moreover, the training of the Bayesian classifier is practical with simple algorithms (Domingos and Pazzani, 1996). This method is remarkably interesting to be applied for recognizing the bolt-position.

This paper presents a bolt-identification algorithm using the signals from low-cost IMU attached on an intelligent screwdriver, to identify which bolt the tool-tip have been arrived and fastened. This algorithm, based on Bayesian decision theory, increases the performance of the tool-tip tracking system by using the identified-bolt position as an input signal to an observation module for an error model based on Kalman filter. Note that lower-case bold letters, Greek or Roman, denote vectors, and upper-case bold letters denote matrices.

2. Bayesian decision theory

Bayesian decision theory is a fundamental statistical approach to the problem of pattern classification (Duda *et al.*, 2001). To illustrate the basic idea of this decision theory, a specific problem, separating salmon and sea bass in a basket is used as an example. When a fisherman picks a fish in the basket, there are two possible states: either the fish is a salmon or the fish is a sea bass. The state of nature is w_1 for picking sea bass, and is w_2 for picking salmon. This state of nature is unpredictable. Therefore, it is described in term of probability.

When there are sea bass as much as salmon in the basket, the probability that the next fish is a sea bass is equal to the probability that the next fish is a salmon. In other words, the priori probability (or simply *priori*) that the next fish is a sea bass $P(w_1)$ is the same as the priori that the next fish is a salmon $P(w_2)$. And the summation of $P(w_1)$ and $P(w_2)$ is one. Assume that the fisherman is forced to make a decision that the next fish is a sea bass or a salmon without being allowed to see. Logically, he follows decision rule: decide w_1 if $P(w_1) > P(w_2)$, otherwise w_2 .

However, in real applications, their decisions are not done based on insufficient information. As the example, the different between these two types of fish is the lightness which is measured as x to improve the fisherman's decision. The lightness of fish x is assumed to be a continuously random variable. Its distribution depends on the state of nature, and is defined as $p(x|w_1)$ for sea bass and $p(x|w_2)$ for salmon. This distribution is called a class-conditional probability density function. Therefore, the difference between $p(x|w_1)$ and $p(x|w_2)$ shows the deference in lightness between populations of sea bass and salmon.

In general cases, the prior probabilities $P(w_j)$ and the conditional densities $p(x|w_j)$ are known, where $j = 1, 2, \dots, c$. The probability of the state of nature being w_j giving that the feature value x has been measured, the posterior probability (or posterior) $P(w_j|x)$ is defined as Bayes' formula in Eq 112.

$$P(w_j|x) = \frac{p(x|w_j) \cdot P(w_j)}{p(x)} \quad [1]$$

where there are c categories in a case:

$$p(x) = \sum_{j=1}^c P(x|w_j) \cdot P(w_j) \quad [2]$$

The likelihood of w_j with respect to x is defined as $p(x|w_j)$, which x is more likely to be the w_j category when $p(x|w_j)$ is larger. Therefore, the evidence factor $p(x)$ is merely a normalize factor to ensure the summation of posterior probabilities is one.

Discriminant function $g_i(\mathbf{x})$ is used for making a decision of classification, where $i = 1, 2, \dots, c$. A feature \mathbf{x} is categorized to class w_i if:

$$g_i g_i(\mathbf{x}) > g_j \text{ for all } j \neq i \tag{3}$$

In other words, it needs to compute c discriminant functions and select the category corresponding to the largest discriminant. The classifier based on this technique is modelled as Figure 1.

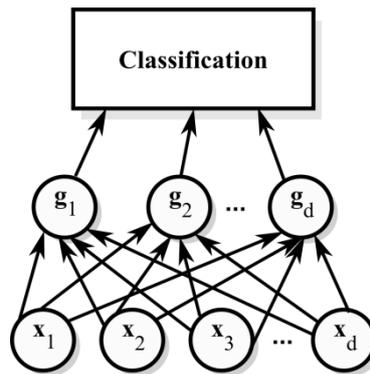


Figure 1: The model of classifier based on discriminant functions

3. Overview of tool-tip tracking algorithms

There are four main algorithm modules as shown in Figure 2, the initialize module, the attitude estimation module, the position estimation module and the bolt-identification module. In brief, the initialize module determines the initial position of the IMU, $r_{ob,0}$, and the initial orientation $q_{n,0}^b$ of the intelligent screwdriver by using the IMU and ultrasound sensors. From the initial position to the first bolt-position, the information of the optimal attitude q_n^b is estimated by the attitude estimation module. Then, the position estimation module determines the optimal position of the tool-tip \hat{r}_{tip} . Finally, the bolt-identification module estimates the recognized position \bar{r}_{reg} of the fastened bolt.

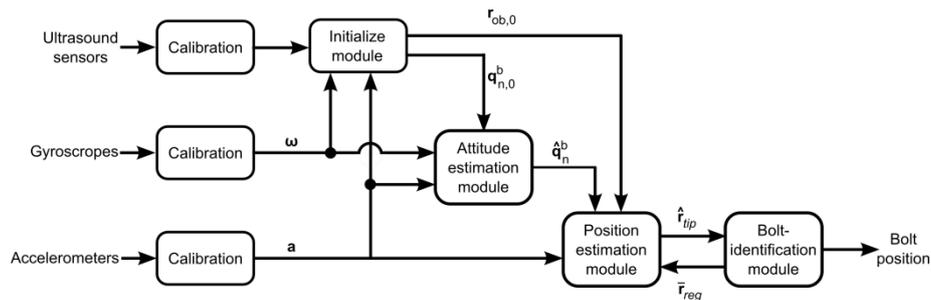


Figure 2: Overview of tool-tip tracking algorithms

4. Design of classification model

4.1. Feature definition

The first step to design a classifier is to decide what feature is suitable for classification. Remarkably, the different among the position of target bolts must be used as one of the feature of this classifier. Because the input of this classifier is the optimal position of the tool-tip \hat{r}_{tip} , there is always the error from the sensor drift in the input signal. In this study, the time length between start and stop moving the intelligent screw driver, λ , is also a feature of the classifier.

$$\mathbf{x} = [\hat{\mathbf{r}}_{\text{tip}} \ \lambda]^T = [\hat{x} \ \hat{y} \ \hat{z} \ \lambda]^T \quad [4]$$

Therefore, the feature \mathbf{x} is defined as Eq 4.1, where \hat{x} , \hat{y} and \hat{z} are the optimal tool-tip position in each axis in initial frame.

4.2. Discriminant function

The second step is to define a discriminant function. As mentioned before, the optimal tool-tip $\hat{\mathbf{r}}_{\text{tip}}$ position always contains the error from sensor drift. Because the position is determined by double integrated the acceleration signals from the IMU, the error from this estimation increased exponentially over time. Therefore, the discriminant function of this classifier should be able to eliminate the effect from the error from sensor drift. In this study, the discriminant function in logarithm form as Eq 4.2 is practically applied for the classification model.

$$g_j = \ln \mathbf{p}(\mathbf{x}|w_j) + \ln \mathbf{P}(w_j) \quad [5]$$

As solved in (Duda et al., 2001), if the densities $\mathbf{p}(\mathbf{x}|w_j)$ is multivariate normal with a d -component mean vector $\boldsymbol{\mu}_j$ and d -by- d covariance matrix $\boldsymbol{\Sigma}_j$, Eq 4.2 can be reformed as Eq [, where d is the number of the components of feature.

$$g_j(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\mathbf{x} - \boldsymbol{\mu}_j) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{\Sigma}_j| + \ln \mathbf{P}(w_j) \quad [6]$$

where

$$\boldsymbol{\mu}_j \equiv \int \mathbf{x}_j \cdot \mathbf{p}(\mathbf{x}) \, d\mathbf{x} \quad [7]$$

$$\boldsymbol{\Sigma} \equiv \int (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \cdot \mathbf{p}(\mathbf{x}) \, d\mathbf{x} \quad [8]$$

Therefore, this classification model will be trained with the measurement data from experiments for estimating a mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$.

4.3. Class definition

The categories w_j of this classifier are depended on target bolts and the initial position of the intelligent screwdriver. Therefore, the number of class, c , is defined as in Eq 4.3, where n is the number of target bolts.

$$c = n + 1 \quad [9]$$

The recognized bolt position, \bar{y} , is the most likely class respect to the feature \mathbf{x} . Moreover, this recognized bolt-position is the output of the bolt-identification module. The recognized bolt-position is selected by comparing the discriminant function as Eq2.

5. Simulation and experiment

To ensure the performance of the bolt-identification module, the algorithm was tested with simulated signals based on Monte Carlo technique (Caflish, 1998). Moreover, the bolt-identification module was integrated in the main module (the tool-tip tracking algorithms) and tested the performance with a test bench.

Based on Monte Carlo method, the simulation was done with 20000 randomized data from two types of simulated signals (1000 data per a bolt-position per signal types). One type of the simulated signals, *SimReq* is a set of randomized features \mathbf{x} which the maximum error of position estimation is within the requirement of the application. Another type is *SimIMU* which is a set of randomized features in range of $\pm 3\delta$, where δ is the standard deviation of estimated position from the previous experiments for testing the position estimation module.

In the experiment, the bolt-identification module is tested with the measurement signal from IMU attached on the intelligent screwdriver as Figure 3. Before this experiment, the parameters of the error model in the attitude estimation module and the position estimation module are already optimized by simulation with white noise.

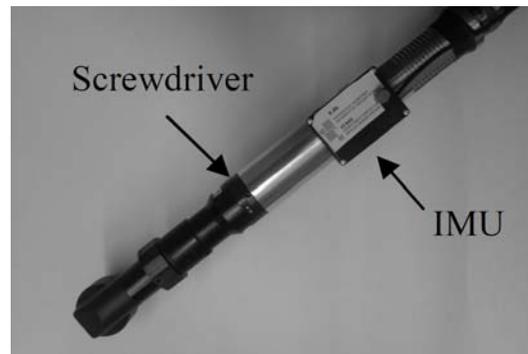


Figure 3: The intelligent screwdriver attached with IMU

To test the performance of the bolt-identification module, the screwdriver is attached with a low-cost IMU. The measurement data from the IMU are acquired by the Labview application which is developed in Competence Center for Applied Sensor Systems (ccass). Moreover, this Labview application determined the arrived bolt position based on the algorithm of the bolt identification module. The interface of the Labview application is shown in Figure 4.

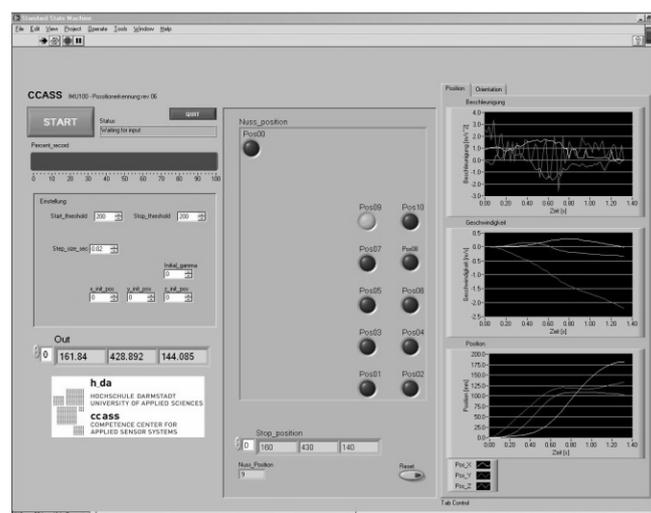


Figure 4: Interface of the Labview application for tool-tip tracking

In this experiment, the measurement data (acceleration and angular velocity signals) are acquired from an IMU100, which is successfully developed by ccass, which is attached on the intelligent screwdriver. The IMU100 is 9-DOF inertial measurement unit integrated with 3-axis magnetometer, a digital thermometer and a digital pressure sensor. The 3-axis accelerometer in IMU 100 is LSM303DLH¹ from ST, which give noise $218 \mu\text{m}/\sqrt{\text{Hz}}$ and its sensitivity is $\pm 20 \text{ mg}$. The gyroscopes of this IMU are LPR430AL² and LY330ALH³ from ST. The rate noise density of LY330ALH⁴ is less than LY330ALH ($0.014 \text{ dps}/\sqrt{\text{Hz}}$ and $0.018 \text{ dps}/\sqrt{\text{Hz}}$ in order). The reference tracking system is the AS200 infrared tracking system from Lukotronic. And the reference position is measured from the infrared markers.

1 http://www.st.com/internet/com/TECHNICAL_RESOURCES/TECHNICAL_LITERATURE/DATASHEET/CD00260288.pdf

2 <http://www.kosmodrom.com.ua/pdf/LPR430AL.pdf>

3 http://www.st.com/internet/com/TECHNICAL_RESOURCES/TECHNICAL_LITERATURE/DATASHEET/CD00254111.pdf

4 http://www.st.com/internet/com/TECHNICAL_RESOURCES/TECHNICAL_LITERATURE/DATASHEET/CD00254111.pdf

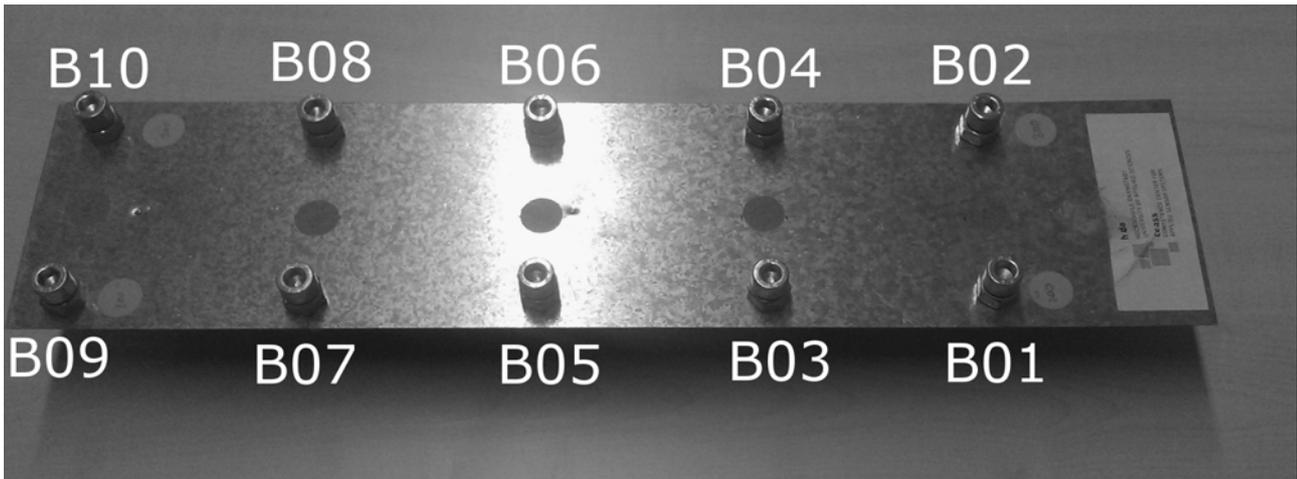


Figure 5: A model of engine cylinder head

The experiment was done by moving the screwdriver from initial position (B0) to a bolt position on a model of an engine cylinder head as in Figure 5 (5 times per bolt position, and 10 target-bolt positions). In detail, the pitch between each bolt position is 10 cm. After processing the measurement data, the estimated positions were compared with the measured attitude from the infrared tracking system.

6. Results

The results from the simulation are shown in Figure 6. Significantly, the bolt-identification module successfully detected the right bolt-positions from the features *SimReq*, which the maximum position error of these features is within the maximum tolerance of the application. Therefore, this module is able to recognize the bolt position as the requirements.

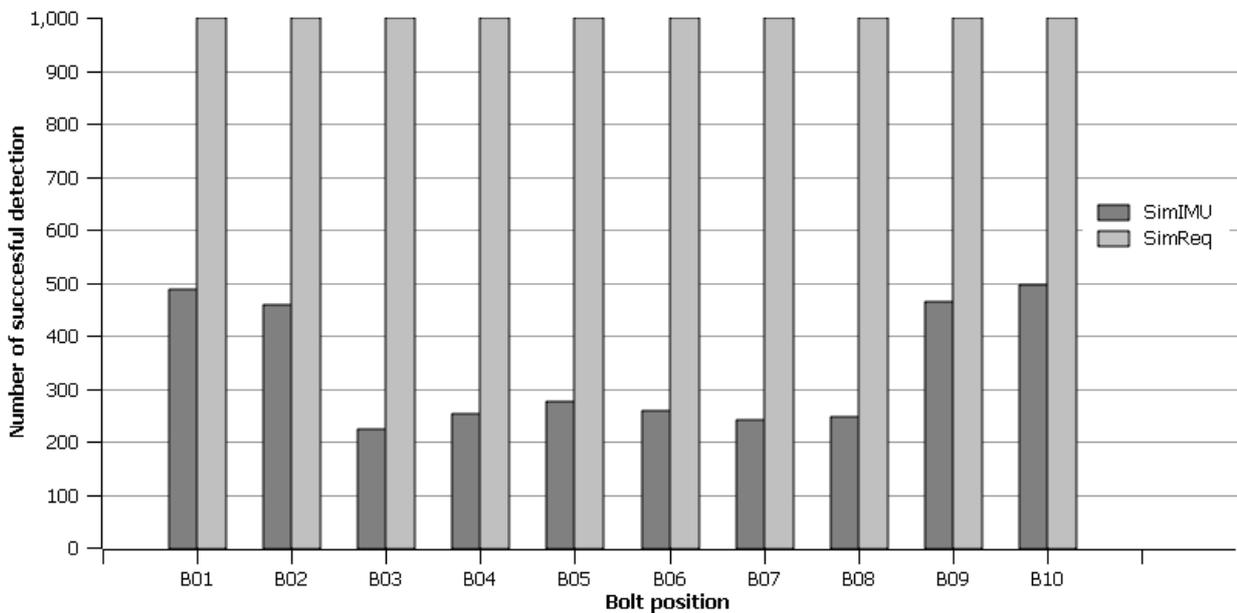


Figure 6: The number of successful detection from the simulation

However, the bolt-identification module poorly detected the right bolt-positions using the features *SimIMU*, which the variance of the data was depended on the standard deviation of the position estimation from the previous experiments. In other words, this poor performance is the result from the position estimation module.

For the experiment, the results of detected bolt positions are categorized in 4 categories. When the detected bolt position is right, the result is counted as "Right" category. "Next" and "Nearby" categories are defined when a detected bolt position are within radius 10 cm. and 15 cm. from the right position in order. A result is in category "Out" when a detected bolt position is not in other categories.

As Figure 7, the performance of the tool-tip algorithms is acceptable, but still it needs to be improved. There is no "Out" category in all tests. Base on the result from the simulation, the Bolt-identification module is successfully recognized a bolt position. However, there is the error from the attitude estimation and position estimation, which decrease the performance of bolt-position detection.

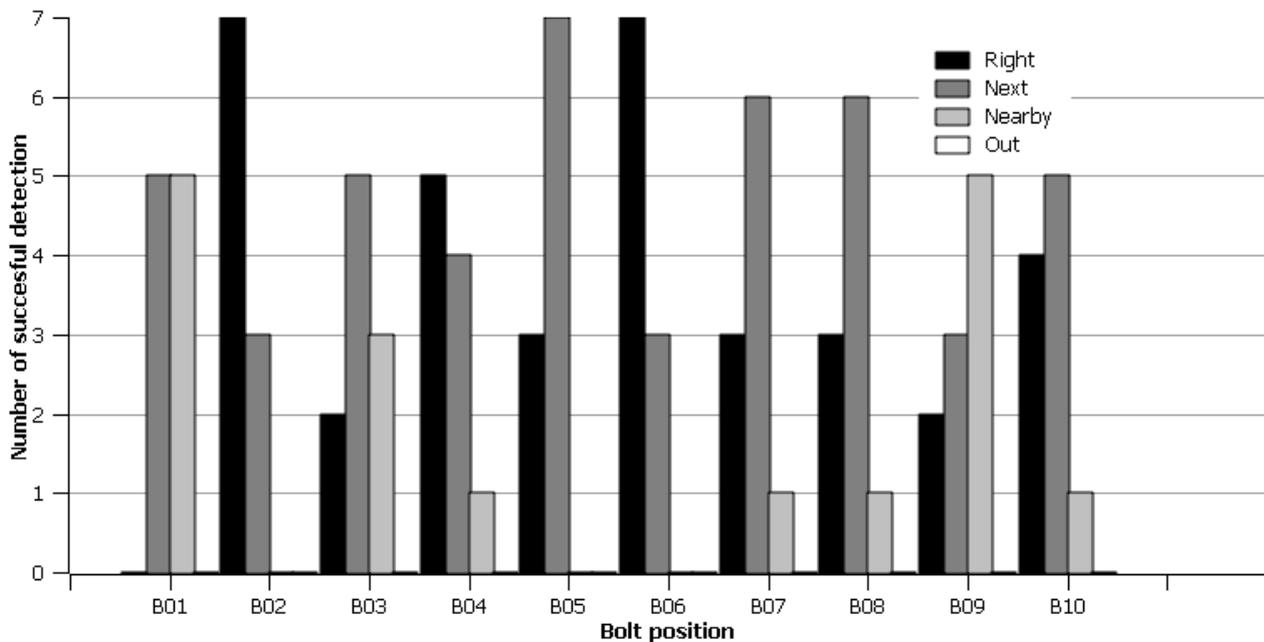


Figure 7: The performance of the tool-tip tracking algorithms

7. Conclusion and Outlook

In this paper, the bolt-identification module using the estimated position from IMU as feature is presented. This algorithm is part of the project "INSCHRAV" that its target is to develop a tracking system to track an intelligent screwdriver in automotive production line, by using low-cost IMU. With the Bayesian decision theory, which is described in this paper, the bolt-position detection is able to be successfully determined as the result from the simulation.

For further development, the attitude estimation module and the position estimation module will be further developed to increase their accuracy for estimation a tool-tip position. The attitude information will be sent to the position estimation module for compensating the effect from gravitation and re-estimating the position of the tool-tip with help of the bolt-identification module. Moreover, this bolt-identification module is able to be implemented in other pattern recognitions, such as auto-focusing for machine interface.

8. References

- Barshan, B. and Durrant-Whyte, H. (1995), "Inertial navigation systems for mobile robots", *Robotics and Automation, IEEE Transactions on*, vol. 11, no. 3, pp. 328–342.
- Caffish, R. E. (1998), "Monte Carlo and quasi-Monte Carlo methods", *Acta Numerica 7. Cambridge University Press*.
- Domingos, P., and Pazzani, M. (1996), "Beyond independence: conditions for optimality of the simple Bayesian classifier", *the Thirteenth International Conference on Machine Learning*.
- Duda, R. O., Hart, P. E., and Strk, D. G. (2001), "Pattern Classification", *Wiley*

Haid, M., Dorf, T., Marquardt, G., and Schneider, U. (2003), "Orientation tracking with low cost sensors as a Kalman filter application", *IAIN 2003, World Congress*, Berlin.

Haid, M. (2004), "Improvement of the referenceless inertial object-tracking for low-cost indoor-navigation by Kalman-filtering (Verbesserung der referenzlosen inertialen objektverfolgung zur low-cost indoor-navigaton durch anwendung der kalman-filterung)", *Ph.D. dissertation, Universität Siegen*.

Hegrenses, O., Hallingstad, O., and Jalving, B. (2007), "Comparison of mathematical models for The Hugin 4500 AUV based on experimental data", *Underwater Technology and Workshop on Scientific Use of Submarine Cables and Related Technologies, 2007*, pp. 558–567.

Langley, P., Iba, W., and Thomas, K. (1992), "An analysis of Bayesian classifiers", *the Tenth National Conference of Artificial Intelligence*.

Titterto, D. H. and Weston, J. L. (2004), "Strapdown inertial navigation technology", *The Institution of Electrical Engineers and The American Institute of Aeronautics and Astronautics*.

Vissiere, D., Martin, A., and Petit, N. (2007), "Using distributed magnetometers to increase IMU-based velocity estimation into perturbed area", *Decision and Control 2007, 46th IEEE Conference*, pp. 4924–4931.

Won, S. H., Golnaraghi, F., and Melek, W. (2009), "A fastening tool tracking system using an IMU and a position sensor with Kalman filters and a fuzzy expert system", *Industrial Electronics, IEEE Transactions on*, vol. 56, no. 5, pp. 1782–1792.

Chapter 5

Electrical Engineering

New approaches for energy optimisation in Smart Homes and Smart Grids by automated service generation and user integration

M. Steinheimer^{1,2}, U. Trick¹, P. Ruhrig¹

¹ Research Group for Telecommunication Networks
University of Applied Sciences Frankfurt/M., Frankfurt/M., Germany

² Centre for Security, Communications and Network Research
University of Plymouth, Plymouth, United Kingdom

e-mail: steinheimer@e-technik.org

Abstract

This paper describes approaches for a framework to design and generate value-added services for management Smart Grids and Smart Homes. It offers users the possibility to design services for managing decentralised energy devices and resources as well as optimise the energy consumption by intelligent energy management. New approaches are introduced that offer automated service generation and optimisation. These new approaches are based on interconnection of households and algorithms for automated optimisation of energy consumption in single households or whole regions, without assistance of third parties. The interconnection of the households rests upon the peer-to-peer principles for communication and automated optimisation as well as forming communities between the participated peers.

Keywords

Smart Grid, Smart Homes, Energy Management, Service Creation, Peer-to-Peer

1. Introduction

Today the power supply system is static and consists of dedicated categories. It has been defined where energy is produced and where energy is supplied to the grid. The energy flow basically takes place only in one way: from power generator in the maximum voltage grid to the energy consumer in the Local Grid. Primary large power plants like coal-fired and nuclear power plants supply companies and households with energy. Energy generation is focussed on central appointments in the power system. The transfer of energy occurs with as much as possible electric tension because of the meanest loss rate. Figure 1 shows the structure of the power supply system today.

The power supply system is separated into four levels of electric tension which are used for transfer and distribution of the energy: Transmission Grid (maximum voltage), Primary Distribution Grid (high voltage), Secondary Distribution Grid (mean voltage), and Local Grid (low voltage). In order to reduce the electric tension to transfer or rather consume the energy on the next lower level, transformers are used at the transition points.

The supply of energy by the central power plants takes place at several levels of electric tension, according to the amount of produced energy. Large power plants like nuclear power plants supply at maximum voltage level, wind parks at high voltage level, small wind power plants at mean voltage grid and sporadically, e.g. photovoltaic power plants at the Local Grid.

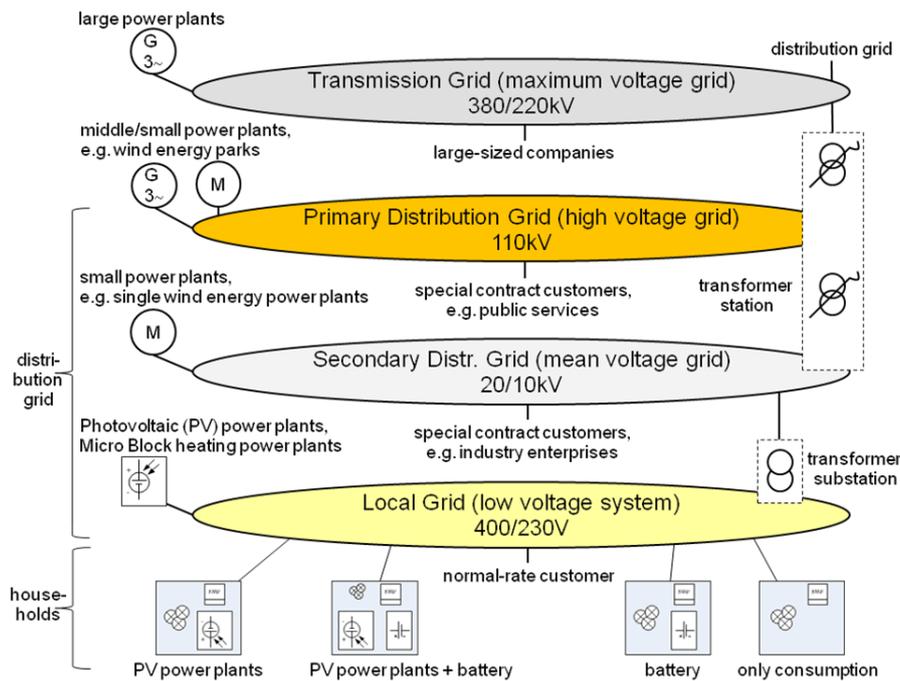


Figure 1: Structure of the power supply system today

The connection of the customers to the different level depends on their demand for energy. Large companies with very high demand for energy are directly connected to the maximum voltage grid. Companies with an average demand are connected to the mean voltage grid and the normal customer is connected to the Local Grid.

In the power supply system a balance of electric current is very important. This implies that the produced and supplied energy has to be equal to the consumed energy. If this is not the case, adjustment control is necessary. Adjustment control today occurs mostly by the central placed power plants. In the Local Grid, where most adjustment is necessary (because the demand for energy is different in specific areas), insufficient possibilities for adjustment exist. The frequency in the grid is one basic indicator to detect the balance of electric current. The 50Hz frequency is measured and corrected on demand. The type of energy used for adjustment is called regulating energy. This energy is very expensive because it has to be provided permanently. Using forecasts allows to detect how much energy will be needed. The more correct the forecast the less adjustment is required and the more convenient is the power production.

In the future the number of decentralised power plants will increase and the central power plants will be reduced. That leads to much more households installing a photovoltaic plant or a micro block heating power plant (BHPP). Furthermore, the number of wind power plants in communes will increase, as well as power generation by biogas. Because the large power plants will be reduced, the still needed adjustment of the power system cannot only occur in the maximum voltage grid. Adjustment in the Distribution Grid is recommended by controlling and monitoring the decentralised power plants and energy consumer. Controlling of these decentralised devices requires fusion of information- and communication technology and energy technology. This will lead to a Smart Grid.

"A Smart Grid is an electricity network that can intelligently integrate the actions of all users connected to it – generators, consumers and those that do both – in order to efficiently deliver sustainable, economic and secure electricity supplies"(GSGF 2012). As a first step towards Smart Grids energy suppliers already install smart meters.

Comprehensive possible solutions for Smart Grids are:

1. The central adjustment (monitoring and control) of distributed generators, storages, and consumers in so-called virtual power plants. This is the preferred solution on the current state of research. In this approach for example, distributed wind power, biogas, photovoltaic plants supplying energy to the primary or secondary distribution Grid can be controlled by the energy provider as well as large energy consumers are switched off for reducing the load. Furthermore, the problem is that the small energy consumers or the power generators and storages in the local grid (households of the customers) cannot be involved in a central control, because the legal basis and technical solutions for controlling of any devices in the user's sovereign territory are missing. Therefore, a more decentralised solution is necessary for optimisation of the energy consumption and energy supply in Local Grids.
2. Decentralised solution for the integration of households in energy management. Here, the energy consumption and the energy purchase are adjusted, controlled and predicted for every household. This approach requires the integration of the customer (in the following "user") and the consideration of their needs. But this is hardly possible because their needs are not known.

One option is energy management or develop/design solutions through value-added services which the user can design active. In addition, the user defines certain boundary conditions and criteria which are used to achieve an added value for the user to optimise energy consumption.

Following approaches will be introduced to allow a control of distributed energy management and monitoring of the power system. With value-added services and the integration of the user in the development and definition of value-added services, a new means of regulating the power grid and controlling the Smart Homes is presented. This approaches will be extended by the automated generation of value-added services and the optimisation of single households. Finally, approaches are presented to achieve optimisation in a Local Grid, through the peer-to-peer internetworking of one or several regions of households (peers) and forming communities between this peers.

2. Services and service platform for energy management

In Smart Home domain some approaches exist to use web services for energy management and connecting heterogeneous systems. (Verschuere et al. 2010, Jisun et al. 2010) adopt service platforms for Smart Homes where services are provided by external service providers. (Shudong et al. 2010) also proposes using web services extended for service orchestration. For fast and easy service creation literature proposes orchestration of web services and composition of services out of reusable building blocks using a service creation environment (SCE). The LOMS-, and the MAMS-project and the open source initiative SPAGIC expose the advantages of graphical service development (Keiser et al. 2008, Freese et al. 2007, Spagic 2012). To solve the permanent rising requirements for services and the underlying heterogeneous communication and execution layer, the projects SeCSE and PLASTIC propose self adaptive service oriented applications (Baresi et al. 2006, Autili et al. 2007).

As described above, several approaches exist in different areas to generate and orchestrate services. Currently, no approach exists where users get the possibility to design personalised services or workflows for energy management in the household, fulfilling their personal needs, including automatic service generation and deployment. Additionally, the approaches are focused on concrete target groups or concrete fields of application, mostly in business field, and are not applicable for users in simple households. Therefore, design and orchestration requires expert knowledge. Currently, there are no known publications that name automated solutions for provisioning of user composed services in Smart Grids and Smart Homes.

Therefore, a continuous solution is needed, oriented at the personal needs of users, to offer cost-efficient energy management, according to user personal needs, and integration within their house automation.

In order to involve users into Smart Grid or Smart Home management a service creation environment (SCE) is needed, which brings the possibility to design and configure value-added services, according to the personal needs of the users, to offer cost efficient energy management and optimised usage of energy networks. A service delivery platform (SDP) is needed to provide automated solutions for service provisioning, service controlling and service management. The following will figure out new approaches for a Service Management Framework (SMF) consisting of SCE and SDP that allows centralised controlling and management of resources and devices in Smart Grids and Smart Homes.

2.1. Energy management in the Local Grid

Monitoring and control respectively adjustment of decentralised power generators and consumers is currently possible on the upper layers of the power grid by the large power plants. Despite the reduction of large power plants still a regulation is required. The control must be in Distribution and Local Grid, where the local power generators supply the current to the power system. The decentralised generators and consumers must therefore be mandatory involved in the adjustment.

Figure 2 shows the increasing decentralised power generators in the Local Grid, and the Service Management Framework which is used for controlling and cross linking of the decentralised generators and consumers in households in the Local Grid as well as decentralised generators in the Distribution Grid.

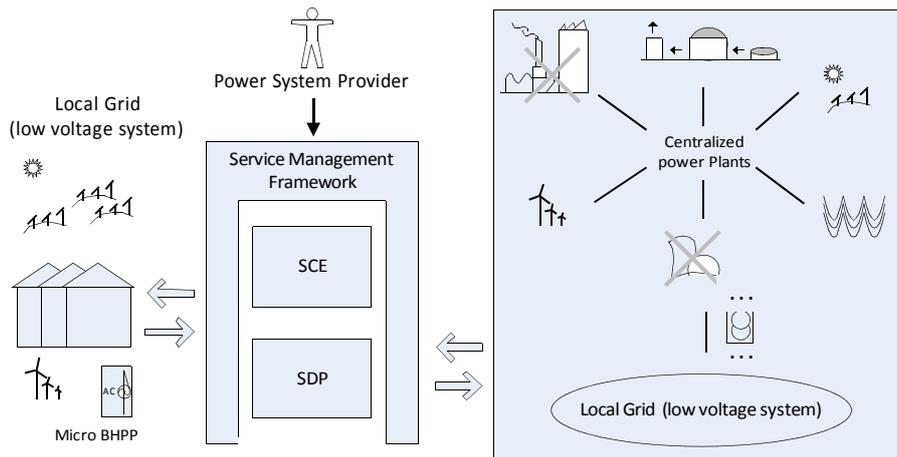


Figure 2: SMF for controlling centralised and decentralised energy resources

The approach through the provision of communication and control services through a Service Management Framework offers power system providers the possibility to implement a central control and monitoring solution in Transmission and Distribution Grids as well as in Local Grids. Furthermore, it is possible to predict the energy demand: whereby the energy operator gets the advantage of providing energy resources timely and with reasonable limits.

As described above, today the consumers, generators and storages in the households cannot be involved in a central control because the legal basis is missing. Therefore in next sections approaches for decentralised solutions are presented by integration of users in the service design.

2.2. Services for Smart Home inclusive energy management

The TeamCom SCE offers the possibility to describe the application flow of a service in a graphical SCE with BPEL (Business Process Execution Language) as service description language. The service described with BPEL will be compiled and deployed on a JAIN SLEE (Java API for Intelligent Network Service Logic Execution Environment) Application Server (Eichelmann et al. 2009, Lasch et al. 2009). This approach comes with the possibility for fast service creation, but recommends knowledge in BPEL and offers no mechanism to integrate these services in Smart Grids or Smart Homes.

The SCE contained in the proposed Service Management Framework makes it possible for the user to design services for its Smart Home in a simple way. Designing services for smart homes requires networking and controllable devices (up to the final consumer). This leads to the connection of electrical devices in households with IT technologies.

Figure 3 shows different energy consumers and energy generators in a smart home, which are connected to each other through various communication networks and controlled by the Service Management Framework.

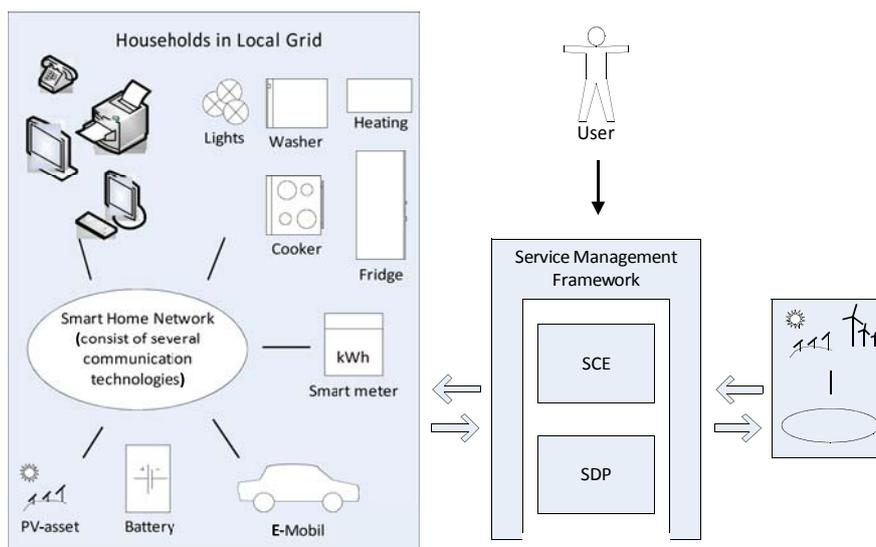


Figure 3: SMF for controlling decentralised energy devices

Power generators and consumers provide device-specific functionality (e.g. timer for start-up and shutdown of equipment or controller for the power supply of power generating facilities). These functionalities can be configured and invoked by the user via the SCE and integrated in the designed service. Generators and consumers are represented graphically in the SCE and form the design interface between the user and the devices. The SCE gives the user the possibility to design own services. In addition, the user can obtain and integrate pre-configured services in the SCE. These services can be offered e.g. as apps from the device manufacturer or a service provider. Thus, for example services can be produced which improve the comfort of the user, reduce the energy costs or increase security. Examples for those services: Increase security by disabling certain devices when the user leaves the apartment (e.g. ensuring that the cooker is switched off). Increase comfort by scheduling starting and stopping of equipment, so they finished their running until defined end points (e.g. washing machine ready when the user comes home). Reduce energy costs by charging the local power storage when energy rate is convenient and using the local power storage as energy resource in inconvenient periods. Also, a service can be designed in order to supply locally produced energy at the best terms to the power system.

In this approach the task of the SCE is to provide an interface for service design and configuration. Additionally, the task contained in the SDP is the execution of these services and the abstraction of communication between the devices in a smart home.

Another function of the SCE is to give users a graphical overview of the energy consumption in its households, the energy consumed by the devices, based on the past and for future planned services. The visualisation of energy consumption also strengthens the awareness of users to use energy more efficiently.

According to the actual technologies, only isolated devices exist which offer the possibility for remote configuration and control. In order to communicate and interconnect devices no unified standard exists. According to the manufacturer devices use proprietary protocols which are incompatible with each other, so networking between devices of different manufacturers currently is hardly possible. To rudimentary activate and deactivate devices, at present approaches are existing using intelligent sockets, which can be predefined to interrupt the power supply. In this new approach, the SMF will offer the possibility for unified communication between the interconnected devices and the integration of device functionalities in user designed services.

2.3. Automated service creation and optimisation

As described above the Service Management Framework will enable the user to create services to control the devices in the Smart Home. This functionality is extended to the automated service creation and optimisation of services and energy consumption. For this, the user can define boundary conditions for energy supply and optimisation, under whose account the SCE generates an optimised service automatically. The user can e.g. define criteria for the price, which advises the SCE not to exceed certain price limit for energy purchases. Or the user defines a lower threshold so that devices only start if the price is below this threshold. Furthermore, users can define criteria for device control and thus define the earliest starting times and latest end times. Similarly, the user can e.g. set the type of power generation for its purchased energy (wind power, hydropower, etc.).

Considering the above-mentioned conditions a service will be designed by the SCE and run by the SDP, which satisfies these criteria and also represents an optimal solution with respect to energy consumption. By scheduling the devices the most continuous energy demand is generated in a household, so that peak loads can be reduced or avoided. An application can e.g. consist of the optimised implementation for the procurement of electricity, where supply, storage and consumption are optimised that the load curve of a household is as flat as possible.

The optimisation of energy consumption by optimal load distribution reduces the load in users' households (scheduling, eliminating temporal peak loads by using batteries for power supply). Forecasts for locally generated and consumed energy calculated by the SCE and sent to the energy vendor enables to keep power plant capacity available within reasonable limits to avoid over- or under-capacity.

Figure 4 shows a standard load profile of an average household (continuous line), and the estimated load curve of a household has been optimised by the SCE (PVU Netze 2012).

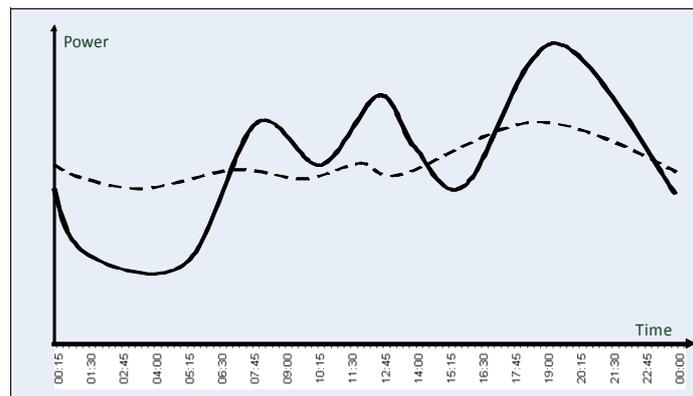


Figure 4: Smoothed load curve in households based on Value-added Services

The maxima, which represent the peak load, can be smoothed considerably, achieved by optimising the load (dashed line). The result is that for these periods no extra power plant capacity will be kept.

3. Peer-to-Peer internetworking and energy community

The approach to optimise a single household is critical because smoothing of the load curve is not sufficient enough to influence the power system without limitation of users comfort. A central control of all households by the energy operator cannot be implemented, caused by legal reasons. The following additional approaches will be introduced to achieve an impact on the power system by simultaneous optimisation of households. Since a central control is not feasible, a decentralised solution is needed.

3.1. Peer to Peer internetworking

A simultaneous optimisation of households requires a network of households. Therefore, communication between the households must exist for exchanging information, which can be used to adjust the energy consumption of a single household on the energy consumption in the region concerned. Figure 5 shows a cross-linking of the households by the Peer-to-Peer principle.

In this approach, the Peer-to-Peer (P2P) networking is chosen because no central authority for control is integrated. Each household will be considered as a peer in the network, which communicates anonymously with other peers. In addition, the other players in the power generation and distribution, e.g. power supplier, distribution system operators, etc., are also considered and integrated as peers in the network.

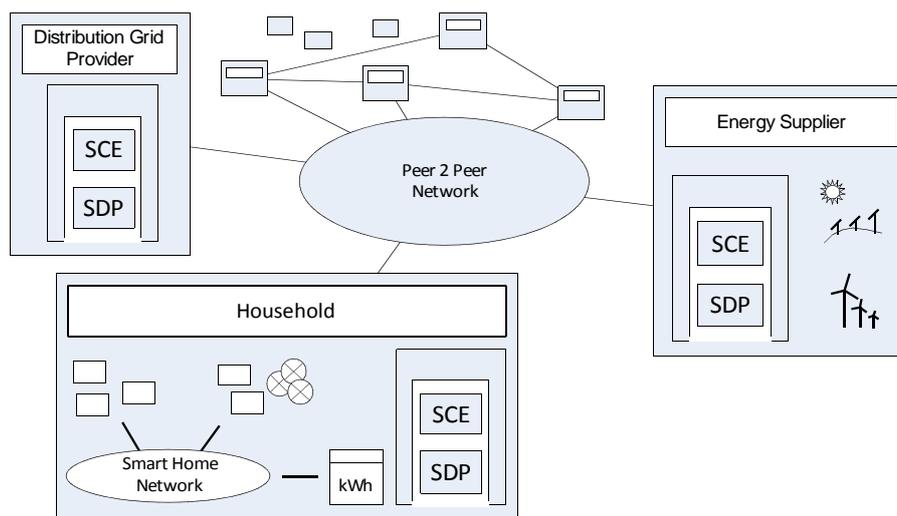


Figure 5: Peer-to-Peer connected households

A self sustained optimisation of every household in a region should lead to an optimisation of the energy requirement in this region. This new approach provides a special algorithm for optimising a single household and the whole region, without assistance of the energy provider, by management and control of the consumers, especially of the generators and energy storages. For the optimisation the exchanged information between peers and the knowledge concerning the own household is used. The exchanged information may include current consumption, forecast of consumption, possible power supply etc.

To ensure the data security and protection of privacy the communication between households has to be anonymous and encrypted. While sharing information between the participating systems, it must be ensured that a third party (e.g. a neighbour) cannot assign this information to a specific household for example, the information about the planned power consumption. To ensure the anonymity, the communication occurs via the Service Management Framework and is not transparent for the users of the respective peers.

In the course of this project, the possible approaches for a P2P internetworking are examined (e.g. Trusted Proxy Provider, Certificate-based P2P networking, etc.). Additionally the different peer-to-peer models for the provisioning of value-added services: Hybrid P2P, Super P2P, Pure P2P (Lehmann et al. 2008) will be analysed for their application in Smart Grids and Smart Homes.

3.2. Energy Community

The approach of the P2P networking will be optimised by extending to the mechanisms of social networks to form a community between households, which have common shared interests and follow the same economic goals. This may include: energy saving, cost reduction, environmental conservation. In a region (e.g. households in a street, city districts) sub-communities can be formed, which appear as part of the whole community. The previously described optimisation of households in this approach occurs between members of the community or within sub-communities. Figure 6 shows the concentration of households in a community consisting of various sub-communities.

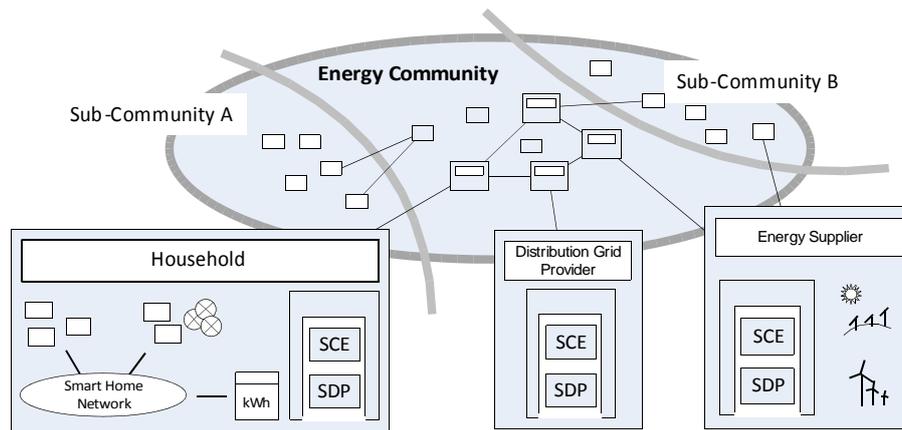


Figure 6: Energy Community with Sub-Communities

According to the principle of a social network, the community may grow independently and every participant may accede to his free will. The possibility to join on a sub-community voluntarily, which matches users' interests promotes the willingness to participate in this system and creates a personal benefit.

Through sub-communities the different interests of the members are considered. Members sharing the same interests come together to achieve their personal goals regardless of other members in the community. Thus, a variety of targets can be achieved e.g. to save energy costs, to reduce CO₂ emissions or to protect the environment. The target may also be to have lower energy consumption than other members of the community. The sub-communities can interact with each other and compare the fuel economy of different regions e.g. challenges may arise with the target to find out which sub-community achieved the lowest energy consumption (or the largest monthly reduction in energy consumption relative to the previous month).

The energy suppliers and power system operators can also occur as part of this community and offer, e.g. for the members of a community, special billing models. Thinkable are flexible tariffs with the possibility to change tariffs and providers dynamically or prepaid billing models. A sub-community can also occur as a service provider for the control area provider by offering adjustment, active and reactive power. Services that are implemented for a sub-community could include the optimisation of power consumption and purchasing. Furthermore, smoothing the load profile for a community which in turn can be honoured by the energy provider with a better deal.

4. Conclusion

This paper presents approaches for a service management framework to control and monitor decentralised energy consumers, storages and generators. In this new approaches the user is integrated in design and configuration of the services for energy management which offer the possibility to follow its personal needs.

The algorithms for automated optimisation and generation of services by controlling decentralised energy consumers, storages and generators, offers the possibility to reduce the energy consumption in households or whole regions. This leads to a possible reduction of costs or environmental benefits. In addition, the power supply system is discharged by reducing load peaks and smoothing the load curve of households and regions. In order to communicate and interconnect the Service Management Framework is used, thus the integration of different devices gets possible und the privacy of involved users is ensured. On the first time approaches point out the peer-to-peer interconnection of households in Smart Grids and build up communities for decentralised optimisation of energy consumption.

In the next research topics algorithms for optimisation of energy consumption between peer-to-peer interconnected households are examined, as well as techniques for abstraction energy devices.

5. References

- Autili, M., Berardinelli, L., Cortellessa, V., Marco, A., Ruscio, D., Inverardi, P., Tivoli, M. (2007), "A Development Process for Self-adapting Service Oriented Applications", proceeding of ICSOC, pp442-448.
- Baresi, L., Di Nitto, E., Ghezzi, C. (2006), "Toward Open-World Software: issues and challenges", IEEE Computer, Volume 39, No. 10, pp36-43.
- Eichelmann, T., Fuhrmann, W., Trick, U., Ghita, B. (2009), "Support of parallel BPEL activities for the TeamCom Service Creation Platform for Next Generation Networks", Proc. of SEIN 2009, pp69-80.
- Freese B., Staiger, U., Stein, H. (2007), "Multi-Access Modular-Services Framework - Whitepaper", Deutsche Telekom Laboratories.
- GSGF Corporation Web Site (2012), "What is the Definition of Smart Grid", <http://www.globalsmartgridfederation.org/smartgriddef.html#Definition>, (Accessed 2 January 2012).
- Jisun, L., Dae-Kyo, J., Yoonkee, K., Young-Woo, L., Young-Myuong, K. (2010), "Smart Grid Solutions, Services, and Business Models Focused on Telco", IEEE/IFIP Network Operations and Management Symposium, pp323-326.
- Keiser, J., Kriengchaiyapruk, T. (2008), "Bringing Creation of Context-Aware Mobile Services to the Masses", IEEE SOA Industry Summit (SOAIS), pp105-106.
- Lasch, R., Ricks, B., Tönjes, R., Lehmann, A., Eichelmann, T., Trick, U. (2009), "TeamCom: A Service Creation Platform for Next Generation Networks", The Fourth International Conference on Internet and Web Applications and Services ICIW 2009.
- Lehmann, A., Fuhrmann, W., Trick, U., Ghita, B. (2008), "New possibilities for the provision of value-added services in SIP-based peer-to-peer networks", Proc. of SEIN 2008, pp167-176.
- PVU Netze Web Site (2012), "Standardlastprofil VDEW", <http://www.pvu-netze.de/media/Standardlastprofil%20VDEW.pdf>, (Accessed 10 January 2012).
- Shudong, C., Lukkien, J., Liang, Z. (2010), "Service-oriented Advanced Metering Infrastructure for Smart Grids", IEEE Power and Energy Engineering Conference (APPEEC), pp1-4.
- Spagic Corporation Web Site (2012), "Spagic, the universal middleware for the development of SOA/BPM extensible solutions and governance of services", <http://www.spagoworld.org/xwiki/bin/view/Spagic/>, (Accessed 20 January 2012).
- Verschuere, T., Haerick, W., Mets, K., Develder, C., De Turck, F., Pollet, T. (2010) "Architecture for smart end-user services in the power grid", IEEE/IFIP Network Operations and Management Symposium, pp316-322.

Simulation and Experimental Validation of a Pseudonoise Method for Multi-Fault Location and Identification in the presence of Noise on Transmission Line Systems

R. A. Guinee, C. Healy

Department of Electronic Engineering
Cork Institute of Technology, Cork, Ireland
e-mail: Richard.Guinee@cit.ie, caroline.healy@mycit.ie

Abstract

This paper demonstrates an innovative pulse tester pseudonoise correlation strategy through numerical simulation and experimental validation in the presence of noise for methodology authentication. This technique employs a Pseudorandom Binary Sequence (PRBS) excitation as a feasible competitive alternative to traditional Time Domain Reflectometry (TDR) in transmission line fault identification and location. System identification resides in the distinctive characteristic signature generated as a consequence of the Cross Correlation (CCR) of both the induced fault echo response and the PRBS input test stimulus. The manifestation of the delayed CCR profile signature also facilitates the attainment of an accurate fault location via the velocity of propagation of the signal over the transmission link. Significant results are presented in this paper demonstrating the methodology's multi-fault detection properties and the effectiveness of multiple PRBS cycle correlation in transmission link noise rejection. Results clearly illustrate noise reduction as correlation cycle number increases ensuring fault resolution and identification is achieved.

Keywords

Autocorrelation, Binary Sequences, Fault Detection and diagnosis, Fault Location, Numerical Simulation, Transmission Line Measurements.

1. Introduction

Transmission line fault analysis and location is a well established technology implemented in the power utilities and telecommunication industries. TDR as the effective testing standard Clegg, B. (1993), Agilent Technologies, (2002), has been fundamental to this discipline for many years in the provision of continuity and security of supply to customers. A continuous online transmission line fault monitoring system with immediate recognition and flagging attributes is crucial to accomplishment of these objectives. TDR operation, however, has a primary shortcoming in that it relies on the principle of a single pulse injection for transmission line fault detection Clegg, B. (1993), Agilent Technologies, (2002). Consequently TDR is susceptible to measurement imprecision from link attenuation with increase in fault distance, phase distortion with frequency and resolution error in the presence of superfluous noise on the link. The alternative proposed strategy presented in this paper encompasses the deployment of a bipolar PRBS input stimulus possessing a randomly arranged time sequence array of low amplitude pulses with a wideband spectrum, Hartley M. G. (1975). This is considered to be white noise like in nature which fundamentally allows system identification, as a fault syndrome eigenfunction extractor, to be undertaken of the transmission link. This methodology is a well recognized system identification tool Godfrey, K. (1993), Guinee R. A. (2000), Horan D. M and Guinee R. A (2005), implemented in control systems design. When the CCR procedure is undertaken with the PRBS test input stimulus and the fault echo response over multiple PRBS cycles, it facilitates the growth of the required CCR fault profile signature by progressively filtering out the surplus noise pick up and normal online signal traffic which consequently results in an accurate fault location and type symptom tag identification. Experimental results are presented for a co-axial cable of fixed line length over a range of known resistive fault terminations ranging from open circuit (O/C) to short circuit (S/C). This paper also exhibits numerical simulation results for a high frequency (HF) co-axial transmission cable model, Microsim pSpice (1997) – RG213/U – which is similar to that used in experimental testing for multiple fault location and identification for a known fixed line length apriori thus further validating the proposed PRBS strategy. Results support the proposed PRBS pulse tester strategy for multiple fault coverage and trace location as both an accurate and viable competitor to TDR for transmission line fault location and diagnosis. The simulation model was also subjected to further testing in the presence of injected Gaussian white noise perturbation, thus emulating normal online cable disturbance or pickup, to illustrate the effectiveness of multiple PRBS cycle correlation for multiple fault location and identification in addition to those presented in, Healy, C. and Guinee, R.A. (2011). Additional estimated fault related quantities such as the reflection coefficient and VSWR derived from obtained experimental results are in close agreement with theoretical calculations thus substantiating the accuracy and reliability of the proposed technique.

2. TDR and PRBS Fault Diagnosis

TDR operates on the theory of a single pulse injection which propagates down the cable until encountering a faulty/mismatched termination. Once the test stimulus encounters this fault/mismatch a full or partial reflection of the pulse energy transpires. This reflected pulse propagates back towards the sending end to the TDR test instrument and thus the fault distance reading is determined. TDR can be deemed an insufficient testing technique as the injected test pulse is susceptible to degradation with phase distortion with transmission distance which inevitably broadens and reduces the sharpness of the pulse as it propagates down the line and also to line attenuation with measurement susceptibility to noise corruption. Pulse resolution is a key feature in determining the fault location in TDR with narrow pulses giving rise to very sharp trace features that are ideal for measurement. However, narrow pulses are easily broadened with increased line distance trouble-shooting because of the frequency response roll-off with signal path transmission distance due to reciprocity of pulse-width with bandwidth Agilent Technologies, (2002). This affects reflected pulse amplitude rectilinearity with shape definition loss and as such narrow pulse TDR is suitable for short distance measurement. Alternatively TDR wide pulse usage produces a wider and more rounded echo trace feature with leading edge transitions that are difficult to gauge and potentially result in an inaccurate fault distance resolution. TDR is also vulnerable to link noise interference which may be present as this can obscure weak long distance fault reflections, Clegg, B. (1993). Hence this can result in pulse definition loss for an accurate fault location measurement. Pulse stimuli can be transmitted repetitively, which are independent and uncorrelated, and the fault information is contained in unconnected pulse echoes rendering repeated pulse injection inefficient in noisy link fault measurements. The PRBS fault location technique presented here employs a random code of bipolar pulses that are reflected by a fault/impedance mismatch as a correlated response build-up over an entire test sequence. CCR processing of the fault echo response with the PRBS test stimulus injection produces a unique CCR profile signature as the fault eigenfunction. By obtaining the time displacement between both the CCR and the autocorrelation (ACR) peaks an accurate estimate of the fault location can be derived once prior knowledge of the signal path propagation velocity is known. The unique CCR amplitude profile yields a characteristic signature facilitating the identification of the fault type present. Link noise presence can be filtered out gradually through a CCR evaluation undertaken over multiple PRBS cycles Guinee R. A. (2000), Geisler, H. and Guinee R.A. (2010), which dramatically reduces the impact of surplus noise by increasing the CCR amplitude-to-noise ratio and thus ‘cleans’ the fault characteristic. This enhances the CCR signature profile for accurate fault identification and location measurement.

3. Cable Modelling for Fault Measurement

A low frequency (LF) co-axial transmission line can be modelled using a lumped parameter ‘T-section’ equivalent circuit as per Figure 1 Fraser W. (1978), Connor F. R. (1972). The characteristic impedance Z_0 and complex propagation coefficient γ of the lumped model can be determined as Fraser W. (1978),

$$Z_0 = \sqrt{(R - jL)/(G - jC)} \quad (1)$$

$$\gamma = \sqrt{(R - jL)(G - jC)} \quad (2)$$

with attenuation and phase change coefficients α and β . The model can be simplified for high frequency (HF) ‘lossless’ operation Fraser W. (1978), by neglecting the distributed resistance R and conductance G in accordance with the condition,

$$L \ll R \text{ \& } C \ll G \quad (3)$$

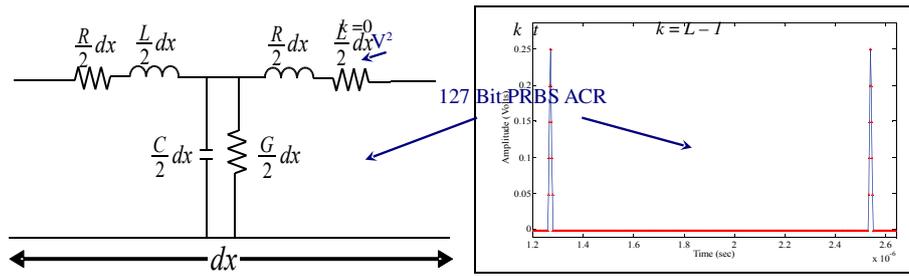
The link propagation velocity can be obtained from the line frequency f and phase-change coefficient β as,

$$v_p = f(2\pi/\beta) \quad (4)$$

The characteristic impedance Z_0 and propagation coefficient γ of a HF ‘lossless line’ with $\alpha = 0$ as per (3) are given by,

$$Z_0 = \sqrt{L/C} \quad (5)$$

$$\gamma = j\omega\sqrt{LC} \quad (6)$$



Figures 1 & 2: Full Model Lumped Parameter ‘T-Section’ & PRBS Autocorrelation Function (ACR)

4. PRBS Line Testing & Fault Location Analysis

The PRBS test stimulant $X(t)$ alternates its logic state pseudo randomly between predetermined voltages $+V$ and $-V$ as a specific coded ‘key’ sequence at discrete time intervals Δt . The test perturbation is generated from a specially configured n stage linear feedback shift register (LFSR) in accordance with a specified ‘key’ polynomial $P(s)$, Hartley M. G. (1975), with a maximum sequence length $L = (2^n - 1)$ and corresponding period $T = L\Delta t$. The PRBS stimulus has a unique delta function like ACR defined for a time shift step $\tau = k\Delta t$ over one period T with $0 \leq k < (L-1)$, as shown in Figure 2 by Guinee R.A. (2009),

$$R_{xx}(k) = \frac{1}{L} \sum_{j=1}^L x(j)x(j-k) = \begin{cases} V^2 & \text{for } k = 0 \\ V^2/L & \text{for } k \neq 0 \end{cases} \quad (7)$$

The time displacement which occurs between the PRBS ACR and the CCR fault response is utilized in this proposed methodology to determine initially the transit delay τ_i and then subsequently the measured fault distance l of the transmission line once knowledge of the propagation velocity v_p has been obtained. In the occurrence of a line mismatch $Z_L \neq Z_0$, the load impedance Z_L is incapable of absorbing all of the incident wave energy and thus a reflection materializes at the load termination with the presence of reflection wave $Y(t)$ along with the incident wave $X(t)$ at any point on the line. The degree of reflection produced by the mismatch is obtained from the load reflection coefficient ρ Fraser W. (1978), and the net subsequent voltage standing wave ratio (VSWR) present is thus deduced from ρ , with $-1 \leq \rho \leq 1$, as,

$$\left. \frac{Y_{rms}}{X_{rms}} \right|_{\text{Termin.}} = \frac{Z_L - Z_0}{Z_L + Z_0} \quad (8)$$

$$s = 1 + \left| \frac{\rho}{1} \right| \quad (9)$$

The various line-fault/mismatched terminations which arise and affect the values of ρ and s Fraser W. (1978), Connor F. R. (1972), can be categorized into the following generalized cases to indicate the line type fault if any that is present Guinee R.A. (2008), Guinee R.A. (2009):

1. Matched load: $Z_L = Z_0$: $\rho = 0$ and $s = 1$.
2. Open circuit (O/C): $Z_L = \infty$: complete incident wave reflection occurs with no phase reversal $\rho = 1$.
3. Short circuit (S/C), $Z_L = 0$: complete incident wave reflection occurs with phase reversal $\rho = -1$.
4. Mismatched termination: $Z_L \neq Z_0$:
If $Z_L < Z_0$ $\rho < 0$ and $s = Z_0/Z_L$ or If $Z_L > Z_0$ $\rho > 0$ and $s = Z_L/Z_0$.

The associated phase relationship between the reflected echo response and the injected PRBS stimulus can be analyzed through ρ in the PRBS – CCR process as well as the VSWR to determine the type of load termination present. A PRBS $X(t) = \{x(1), x(2), \dots, x(L)\}$ when injected into a faulty transmission line generates a conditioned echo response as $Y(t) = \{y(1), y(2), \dots, y(L)\}$ which when cross correlated with the incident disturbance $X(t)$ as Guinee R.A. (2008),

$$R_{xy}(k) = \frac{1}{L} \sum_{i=1}^L x(i)y(i-k) \quad k = 0, 1, \dots, L-1 \quad (10)$$

yields a unique characteristic CCR line fault profile signature. A CCR fault related correlation peak occurs at a specific time shift τ_i in relation to the PRBS ACR. The time displacement τ_i of CCR peak is thus measured from the incident PRBS ACR reference peak, corresponding to time shift $k = 0$ in the CCR process, as illustrated in Figures 3 and 4 for

open circuit and short circuit faults respectively Guinee R.A. (2008). The measurement of τ_l is then divided by two and multiplied by the line propagation velocity v_p to estimate the fault distance l as,

$$l = v_p (\tau_l / 2) \tag{11}$$

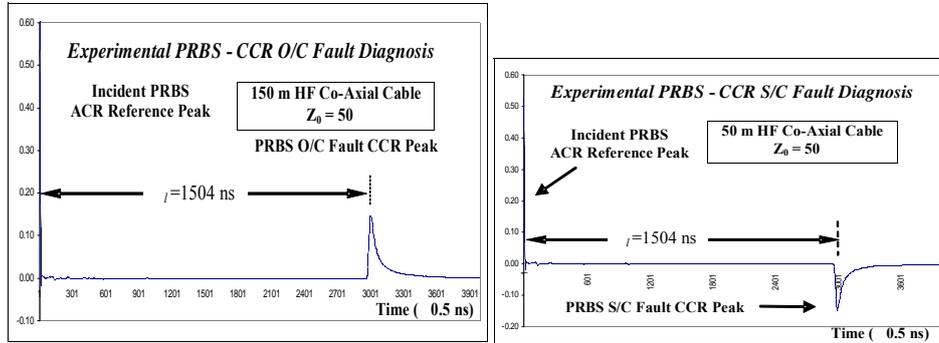


Figure 3 & 4: Experimentally derived O/C & S/C Fault Diagnosis

5. Coaxial Cable Experimental Test Results

A 500 m drum of URM-43 HF co-axial cable with a characteristic impedance $Z_0=50\Omega$ and a distributed capacitance $C=100$ pF/m was cut to a length of 150 m with a remaining drum residue of 350 m Guinee R. A. (2008), Guinee R.A. (2008), Guinee R.A. (2009), Under laboratory controlled conditions the cable test length was then subjected to a PRBS test stimulus injection excitation for various ‘known’ apriori faulty/mismatched terminations, to validate experimentally both the fault location and identification capability using the PRBS generated CCR fault envelope profile. Assuming ‘lossless’ HF line behavior as in (3) and a PRBS test frequency of a 100 MHz the distributed inductance (L) can be derived, for fault location estimation during testing, as,

$$L = CZ_0^2 = 0.25 \text{ H/m} \tag{12}$$

and the subsequent line propagation velocity as $v_p = 1/\sqrt{LC}$ with estimate,

$$v_p = 1/\sqrt{(0.25 \times 10^{-6})(100 \times 10^{-12})} = 2 \times 10^8 \text{ m/s} \tag{13}$$

The estimated velocity of propagation of the line v_p can now be implemented with the echo fault response transit time to determine the fault location as per (11).

5.1. PRBS Characteristics and Fault Distance Resolution

A 4 volt amplitude, 1023 bit PRBS input excitation was utilized as the simulated injected test stimulus with a 10 ns bit duration Δt . The selected Δt corresponds to the co-axial cable specified operational frequency of 100 MHz which was also within the test pattern generator (TPG) limits. Implementing a bit duration of $\Delta t = 10$ ns with a line velocity $v_p = 2 \times 10^8$ m/s, the fault distance can be resolved down to an accuracy Δd given by $\Delta d = (v_p)(\Delta t) = 2$ m. An 8-channel Agilent mixed storage oscilloscope with a 2 GHz sampling frequency was utilized at the line input end in order to capture data of the input stimulus as well as the delayed fault echo response for subsequent post test data analysis and cross correlation signal processing. The higher sampling frequency results in an improved fault distance resolution accuracy of $\Delta l = v_p \Delta t / 20 = 0.1$ m.

5.2. Experimental Co-Axial Cable Test Results

Initially the 150 m cable length was terminated for s/c and o/c conditions to gauge the accuracy of the PRBS - CCR method of fault identification and location as shown in Figures 3 and 4 respectively. Thus the location l of the O/C and S/C terminations is determined from the displacement $\tau_l = 1504$ ns of the CCR fault response from the ACF reference peak via (11) as $l = (2 \times 10^8 \text{ m/s}) (1504 \times 10^{-9} \text{ s}) / 2 = 150.5$ m which is equivalent to the actual physical cable length. Additional line testing was then undertaken with ‘known’ resistive fault terminations $Z_L \neq Z_0$, in 10Ω steps, beginning at 10Ω up to 100Ω and thereafter in 100Ω steps as per Figure 5. Each CCR peak transpires at the same time displacement in Figure 5 as testing was undertaken on a fixed cable length of 150 meters. An important observation can be made from the cable experimental test results in Figure 5 regarding the nature of the fault/mismatch present through

the CCR peaks and their corresponding polarity. Initially an inspection can be made as to whether or not a CCR peak is in fact present. In the event of an absence of a CCR peak, this implies that no mismatch/fault has transpired and thus $Z_L = Z_0$. However if $Z_L > Z_0$ a positive peak is apparent and a potential O/C fault or high impedance mismatch has occurred on the line as seen in Figure 5. Alternatively if $Z_L < Z_0$ then a negative peak is present indicating a potential S/C fault occurrence or low impedance mismatch on the line as per Figure 5. It can be seen that the CCR peak magnitude fluctuates with load Z_L and the extent of mismatch with Z_0 is manifested in the reflection coefficient ρ in (9) which is proportionally passed to the correlation peaks. The CCR peak magnitude increases with fault termination R_L as per Figure 5, with a corresponding increase in the reflection coefficient ρ as per Figure 6, such that,

$$\max R_{xy}|_{R_{L2}} \quad \max R_{xy}|_{R_{L1}} \tag{14}$$

for termination conditions $R_{L2} > R_{L1} > Z_0$. Similar conclusions prevail for the converse case in Fig. 5, which depicts a negative polarity with $R_L < Z_0$ and increased absolute CCR value with reduced fault resistance termination. PRBS fault diagnosis can also be used to estimate the fault echo response reflection coefficient ρ for a ‘lossless’ line for a given fault termination condition R_L as Guinee R.A. (2008), Guinee R.A. (2009),

$$\hat{\rho} = \frac{\max R_{xy}|_{CCR}}{\max R_{xy}|_{ACR}} \tag{15}$$

from a ratio comparison of the ‘reflected’ CCR to the incident ACF peaks in (15) for each of the resistive fault cases R_L in Figure 5. For a co-axial cable with finite losses, the reflection coefficient estimate shown in Figure 6 is determined from apriori knowledge of the O/C or S/C CCR value as Guinee R.A. (2009),

$$\hat{\rho} = \frac{\max R_{xy}|_{CCR}}{\max R_{xy}|_{CCR \text{ O/C}}} \quad \frac{\max R_{xy}|_{CCR}}{\max R_{xy}|_{CCR \text{ S/C}}} \tag{16}$$

This information can then be used to estimate the actual resistive fault manifestation R_L from the expression Guinee R.A. (2008), as

$$\hat{R}_L = (1 - \hat{\rho}^2)^{-1/2} Z_0 \tag{17}$$

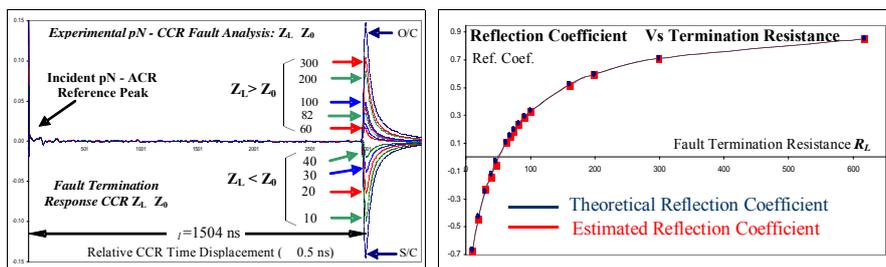
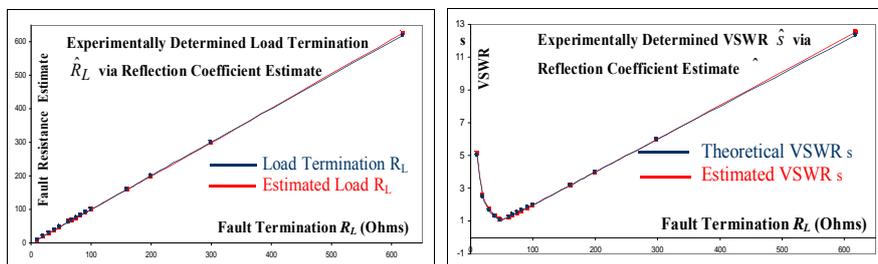


Figure 5 & 6: Experimental CCR Fault Responses for Terminations $Z_L \neq Z_0$ & Reflection Coefficient Experimental Estimates



Figures 7 & 8: Plot of Load Terminations Experimental Estimates & Experimental VSWR Estimates

and the VSWR s , from (9), as,

$$\hat{s} = 1 + \left| \frac{\hat{\rho}}{1 - \hat{\rho}} \right| \tag{18}$$

which are both illustrated in Figures 7 and 8 respectively. Comparison of the derived fault resistance \hat{R}_L and VSWR \hat{s} estimates from the experimental reflection coefficients $\hat{\rho}$ in Fig. 6, with the known resistive terminations R_L used and

theoretical VSWR values illustrate a good correlation when plotted in Figures 7 and 8 respectively. The accuracy of these derived quantities further enhances confidence in the PRBS fault diagnostic method and confirms its capability in fault resistance identification.

6. Simulated HF Line and Multi-fault Diagnosis

A 50m length of HF co-axial cable model was subjected to simulation testing through a ‘pSpice’ circuit simulator, Microsim pSpice (1997) in accordance with (3) for a ‘lossless line’ with $\alpha = 0$. The cable model RG213/U, which is practically identical to the URM-43 HF co-axial cable used in experimental testing has the following parameters $Z_0 = 50 \Omega$ and nominal capacitance $C = 100 \text{ pF/m}$. The propagation velocity of the transmission link can thus be determined as,

$$v_p = \frac{1}{\sqrt{LC}} = \frac{1}{Z_0 C} = \frac{1}{(50)(100 \times 10^{-12})} = 2 \times 10^8 \text{ m/s} \quad (19)$$

The simulation model was injected with a 127 bit PRBS test stimulus for multiple fault identification and location. A 0.5 volt peak-to-peak bipolar (-0.5V to +0.5V) amplitude PRBS was employed with a 10 ns bit duration corresponding to the specified operational frequency of 100 MHz. Each bit duration was sampled 5 times producing an improved bit resolution of 2ns and when combined with the known line velocity of $v_p = 2 \times 10^8 \text{ m/s}$ the fault distance can be resolved to an accuracy Δd given by $\Delta d = (v_p \text{ m/s}) (\Delta t / 5) = 0.4 \text{ m}$. The model simulates two mismatched terminations to gauge the accuracy of and further validate the proposed methodology. The faulty terminations occur at 30 and 50 meters from the source with corresponding mismatches of 45Ω and 200Ω respectively. The injected test stimulus is propagated down the line until encountering the first mismatched termination of 45Ω where a partial reflection back to the source with a phase reversal of the test stimulus occurs. Consequently a negative CCR peak is evident indicating $Z_L < Z_0$ as in Figure 9. As only a partial reflection occurs at the first termination a fractional quantity of the test stimulus continues to propagate down the line to detect the second mismatched termination of 200Ω . On encountering the second mismatched termination, a partial reflection again transpires and the test stimulus is again reflected back to the source with no phase reversal as $Z_L = 200 \Omega > Z_0$. Hence a positive CCR peak is apparent indicating that $Z_L > Z_0$ as illustrated in Figure 9. The measured time displacement $\tau_1 = 304 \text{ ns}$ in Figure 9 between the first CCR peak and the ACF peak results from the propagation delay of the PRBS stimulus to traverse the line from the input to the fault and back with a total distance coverage $2l_1$. Thus the measured time displacement τ_1 obtained in Figure 9 along with the known phase velocity v_p in (19) provides an accurate estimate of the first and second mismatched locations \hat{l}_1, \hat{l}_2 and their type via the reflection coefficient can be determined with the measured time displacement $\tau = 504 \text{ ns}$ as,

$$\hat{l}_1 = \tau_1 / 2 * v_p = (304 \text{ ns} / 2) * 2 * 10^8 = 30.4 \text{ m} \quad (20)$$

$$\hat{l}_2 = \tau_2 / 2 * v_p = (504 \text{ ns} / 2) * 2 * 10^8 = 50.4 \text{ m} \quad (21)$$

Once the CCR peak time displacement and the link propagation velocity v_p is established an accurate estimation of the fault location can be obtained and identified.

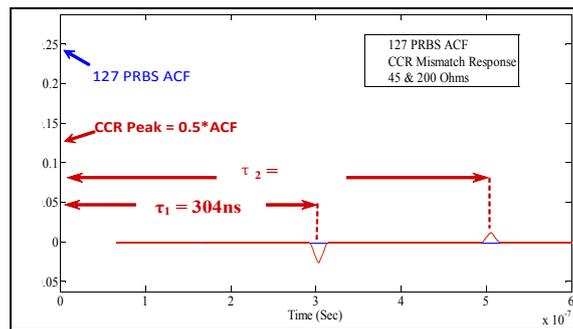


Figure 9: Simulated CCR Signature for 45 & 200 Ohm Mismatched Terminations for a 50m HF (RG213/U) Co-axial Cable.

7. Simulated HF Line and Multi-fault Diagnosis

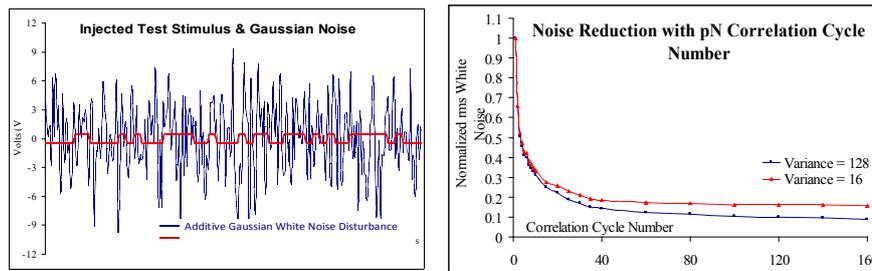
In the majority of realistic fault detection scenarios natural occurring noise exists on transmission links as a result of inductive coupling with electromagnetic interference (EMI) switching noise-like sources. The manifestation of this link noise has significant adverse effects for TDR long distance measurements as a result of its poor echo pulse energy-to-noise ratio response. The proposed methodology of a multi-cycle PRBS injection test stimulus during fault testing and the resulting cross correlated accumulated response rejects the presence of this link noise thereby accentuating the resultant peak CCR amplitude-to-noise ratio. This enables fault termination identification through pickup of weak fault reflections with screening from noise. Significant simulation studies demonstrating multi-cycle PRBS noise rejection capability were undertaken using the co-axial cable model setup as per Fig. 10. This encompassed two noisy test case scenarios using zero mean ($\mu = 0$) additive Gaussian white noise (AGWN) sources with variances $\sigma^2 = 128$ and 256 which were coupled to the 30m co-axial cable segment at the sending ending along with the PRBS test signal thus presenting maximum disturbance to the injected binary test stimulus. Figure 11 illustrates the waveforms of the injected PRBS test stimulus along with the AGWN disturbance, with zero mean and variance $\sigma^2 = 128$, which was then used in the cross correlation of the reflected echo response mismatch over multiple pseudonoise cycles for uncorrelated noise disturbance rejection. One hundred and sixty PRBS test stimulus cycles characterised by a 0.5 volt bipolar amplitude were applied to the noisy 50 Ω coaxial cable input with a 45 Ω and 200 Ω shunt load termination at 30m and 50m respectively. Figure 12 shows the plot in the reduction of the normalized rms noise residue contamination in the CCR fault signature resolution for both AGWN simulation test cases with PRBS cycle correlation number. The error produced between the multi-cycle correlated noisy CCR response and its noiseless equivalent, clearly illustrates the improved visibility and resolution accuracy of the desired CCR fault signature with increased correlation cycle number. A significant observation can be made from Figure 12, regarding the noise related error reduction, it that it can be readily seen that the vast majority of the noise 'filtering' process is undertaken over the initial ten cycles of the CCR process. It can also be observed from Figure 12 that the more pronounced the AGWN present on the link is the more effective the PRBS methodology works in filtering out this undesired disturbance.

<i>PRBS Test Sequence:</i>	Seq-Length L=127 Bits Amplitude: 0.5 volts
<i>Gaussian White Noise Generator:</i>	Peak Amplitude A: 0.5 volts Mean = 0; Variance $\sigma^2 = 256$ & 512

Figure 10: Co-axial Cable Model with Gaussian Noise Injection

8. Conclusion

In this paper a novel PRBS strategy for fault tracing and identification on a HF co-axial transmission line has been validated through both simulation and experimental testing as a competitive alternative to the standard TDR approach. This novel trouble-shooting mechanism relies upon the random attributes of maximal length PRBS sequences and their distinctive delta/spike-like ACR for faultfinding. Hence it can be deployed in the impulse response estimation of a faulty transmission line, in order to identify the fault type and its location, through cross correlation of the reflected response with input PRBS test stimulus. The line fault echo signature is the magnified collective response to a time arranged PRBS sequence of random pulse stimuli, propagated down the line towards the fault termination, and as such is the main advantage of using PRBS testing in preference to TDR. This novel test strategy has been validated for HF co-axial transmission lines for a range of fault impedance terminations ranging from open to short circuit types. The method has been further validated through accurate extraction of the reflection coefficient and VSWR estimates from experimental testing and their comparison with derived theoretical estimates for various known fault terminations. Furthermore the capability of this pseudonoise test technique in multiple fault location and identification has also been demonstrated from simulation. CCR fault syndrome results have been presented using simulation to demonstrate the inherent noise rejection capability of multi cycle PRBS correlation. This is effective for long distance weak fault signature extraction from noisy transmission links and as such demonstrates the efficacy and feasibility of the PRBS test method.



Figures 11 & 12: 127 Bit PRBS Test Stimulus in presence Gaussian Noise & Simulated Noise Roll off with increase in PRBS Correlation Cycle Number.

9. References

- Clegg, B. (1993), "Underground Cable Fault Location", *McGraw Hill*.
- Agilent Technologies, (2002) "Time Domain Reflectometry Theory", Application Note 1304-2, www.agilent.com.
- Godfrey, K. (1993), "Perturbation Signals for system Identification", *PHI, New York*.
- Guinee R. A. (2000), "Variable Speed Motor Drive Testing and Parameter Identification using Pseudorandom Binary Sequences", *IEEE-Melecon, Cyprus*.
- Horan D. M and Guinee R. A (2005), "A Novel Pulse Echo Correlation Tool for transmission Path Testing and Fault Finding using Pseudorandom Binary Sequences", *IEEE International Symposium on defect and fault Tolerance in VLSI Systems*, pp. 229-237.
- Microsim pSpice (1997), "Circuit Analysis Software", *Ver.8.0, Microsim Corp*.
- Fraser W. (1978), "Telecommunications", *Macdonald and Jane's Publishers, London*.
- Connor F. R. (1972), "Waves", *Arnold publishers, Great Britain*.
- Hartley M. G. (1975), "Digital Simulation Method", *Peter Peregrinus Ltd., England*.
- Cooper G. R. and McGillem C. D. (1986), "Modern Communications and Spread Spectrum", *McGraw Hill, NY*.
- Geisler, H. and Guinee R.A. (2010), "A novel high frequency pseudo noise correlator hardware design for cable fault diagnoses", *IEEE PRIME, Berlin*
- Guinee R. A. (2008), "A novel correlation pulse echo methodology for transmission line fault identification and location using pseudorandom binary sequences", *MMACTEE'08, Proceedings of the 10th WSEAS International Conference on Mathematical Methods and Computational Techniques in Electrical Engineering, Bulgaria*.
- Guinee R.A. (2008), "A novel pulse echo correlation tester for transmission line fault location and identification using pseudorandom binary sequences", *IEEE - IECON, Orlando, Florida*.
- Guinee R.A. (2009), "A Novel Pulse Echo Correlation Tester for Transmission Line Fault Location and Identification using Pseudorandom Binary Sequences", www.tyndall.ie/nap/Guinee.pdf, Tyndall National Institute, Cork, Ireland.
- Healy, C. and Guinee, R.A. (2011), "Numerical Simulation and Experimental Validation of a Transmission Cable Statistical Fault Identification Method", *CERC2011 - Collaborative European Research Conference, Cork, Ireland*.

Chapter 6

Civil Engineering

Response of Precast Prestressed Concrete Circular Tanks Retaining Heated Liquids

M. J. Minehane¹, B. D. O'Rourke²

¹ RPS Group Ltd., Innishmore, Ballincollig, Cork, Ireland

² Department of Civil, Structural & Environmental Engineering
Cork Institute of Technology, Bishopstown, Cork, Ireland

e-mail: michael.minehane@rpsgroup.com

Abstract

The present study investigated the influence of heated liquid storage, upward to 95 °C, on precast prestressed concrete circular tanks. Modern design standards for liquid-retaining concrete structures require that thermal effects be considered for the serviceability limit state. They are also to be considered for the ultimate limit state when deemed significant. The standards, however, do not provide guidance for the analysis of such effects. Research in this area is also limited and almost exclusively concerned with ambient thermal conditions, with a maximum temperature change of 30 °C in any instance.

A finite element study incorporating thermo-mechanical coupling investigated the magnitude of stresses associated with thermal storage. A linear eigenvalue analysis revealed that for the tank walls, the ultimate limit state of buckling is a necessary consideration due to the thermally induced combined axial compression and bending. Consequent design implications were established and recommendations made for accommodating thermal loading.

Keywords

Finite element analysis, Prestressed concrete, Reservoirs, Thermal storage

1. Introduction

Precast prestressed concrete circular tanks typify the most efficient form of storage for fluid and granular material. These types of tanks have progressively gained popularity due to their enhanced durability, high quality control, low initial cost and minimal maintenance compared to a steel alternative. Modern precast concrete tanks are commonly proprietary systems comprising wall panels that are prestressed vertically and post-tensioned circumferentially. The panels are generally permitted to slide freely during stressing to enable the entire height of wall to receive the desired residual compression. A ring beam is subsequently cast at the base, thereby providing a pinned footing in service and ensuring liquid-tightness. Figure 1 shows a typical tendon prestressed precast concrete tank.

The significance of thermal effects on concrete reservoir walls for ambient conditions is long established. An early study by Priestley (1976) determined that tensile stresses large enough to cause cracking can occur when the effect of solar radiation is considered. Ghali and Elliott (1992) developed closed form solutions for the thermal analysis of tank walls of varying base restraint and free at the top. Through numerical examples, it was shown that a gradient of 30 °C through the wall is sufficient to cause cracking. This advocated Priestley's (1976) proposal that the design should be based upon a serviceability criterion of limiting crack widths rather than a limiting tensile stress. Although modern design standards require that thermal effects are considered for the serviceability limit state, few provide guidance for the analysis of such effects. Pioneering design codes with regard to this are NZS 3106 (1986) and AS 3735 (1991) which provide design tables, originally derived by Priestley (1976). The design aids are applicable to tank walls free at the top and either free-sliding, pinned or fixed at the base.

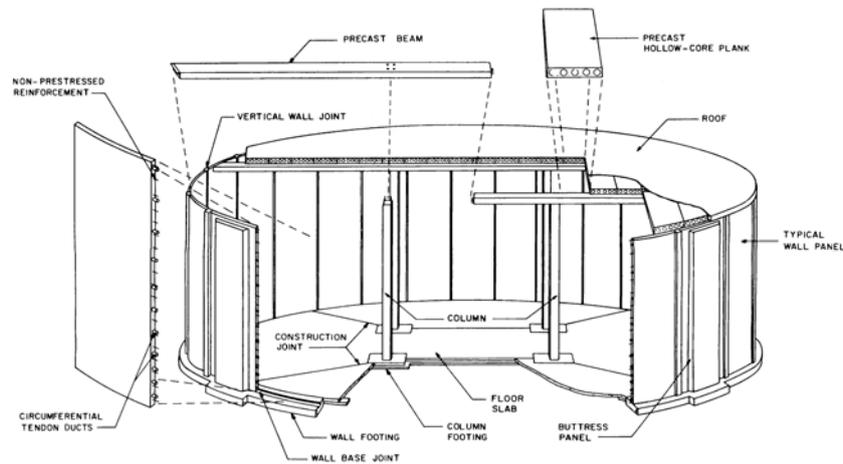


Figure 1: Typical precast prestressed concrete tank, Tadros and Joo (1987)

The studies reviewed were exclusively applicable to tank walls free at the top. As thermal storage tanks require a roof, the associated radial restraint at the top of the wall alters the internal force distribution. Moreover, the magnitude of the internal forces resulting from the thermal expansion of the tank walls will be shown to be prohibitive from a design perspective, unless allowances are made for radial displacement during service.

2. Research significance

This paper investigates the feasibility and implications of thermal storage using cylindrical concrete reservoirs, for which there is currently a paucity of information. The research has practical applications in the oil and gas and nuclear containment industries, in addition to thermal storage for district heating and related schemes. Although particular reference is made throughout to precast prestressed concrete storage tanks, the research is also applicable to partially prestressed and reinforced concrete reservoirs.

3. Influence of elevated temperatures on material properties

EN 1992-1-2 (2004), EN 1992-3 (2006) and FIB Bulletin 55: Model Code 2010 define the reductions in the mechanical properties of both the concrete and steel reinforcement for elevated temperatures. For the temperature range under consideration, the associated strength reductions are insignificant.

Creep of concrete increases with increasing temperature. Fortunately, there has been extensive research carried out on the influence of temperature on concrete creep for structures used in nuclear containment. It would appear from the literature that the use of a thermal scaling factor, or creep coefficient multiplier, is appropriate in accounting for temperature effects on creep. Figure 2 presents a comparison of creep coefficient multipliers from guidance provided by CEB 208 (1991) and FIB Bulletin 55: Model Code 2010 in addition to experimental studies carried out by Brown (1968), Gross (1975) and Nasser and Neville (1965). For a temperature of 95 °C, the creep coefficient multipliers range from approximately 1.95 to 2.44.

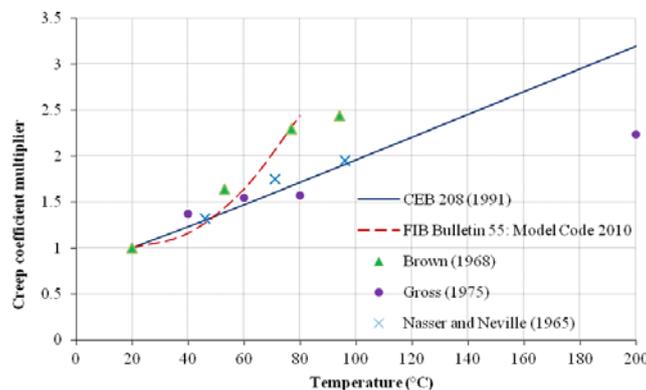


Figure 2: Creep coefficient multipliers with increasing temperature

An accurate evaluation of creep at elevated temperatures is difficult to attain, as creep is sensitive to the evaporable water in the mix. Consequently, an accurate value of the moisture content is desirable if a precise assessment is to be made. The moisture content, particularly at elevated temperatures, is sensitive to the member thickness. The majority of the experimental results were developed for the walls of nuclear containment structures, which are generally deep members comprising wall thicknesses of 1000-1500 mm. Since the walls of precast storage tanks are much thinner, typically 150-200 mm, the walls would lose much more moisture comparably, which would suggest a lower value of a creep coefficient multiplier would be appropriate. The experimental results are generally based upon uniaxial tests. As prestressed concrete circular tanks are subject to a multiaxial state of stress, a reduction in the predicted creep strain would be appropriate, in line with the findings of Hannant (1968) and McDonald (1978).

Numerous experimental studies conclusively reveal that bond strength decreases with increasing temperature. Figure 3 compares results of studies carried out by Harada *et al.* (1972), Kagami (1975), Haddad *et al.* (2008), Chiang and Tsai (2003), Bazant and Kaplan (1996) and Huang (2010).

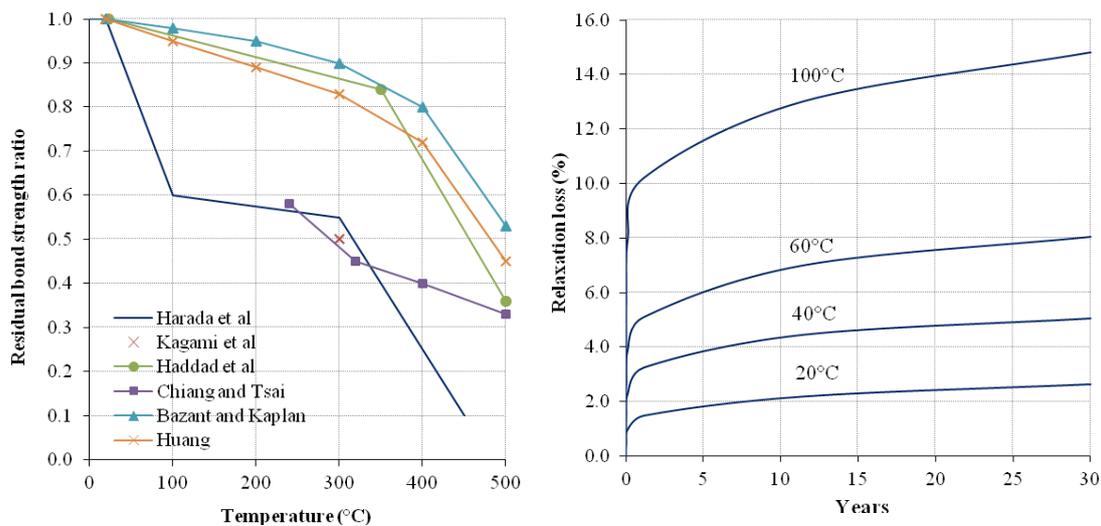


Figure 3: Residual bond strength ratio and stress relaxation loss with increasing temperature

Owing to different coefficients of thermal expansion, steel expands relative to concrete with increasing temperature. Consequently, for pretensioned members, this will effectively increase the loss of prestress due to stress relaxation. FIB Bulletin 55: Model Code 2010 quantifies the increase in loss due to stress relaxation with increasing temperature for a duration of 30 years (Figure 3). The significance of high temperatures on the stress relaxation is evident as a value of approximately 2.5% at ambient increases to approximately 15.0% at a temperature of 100 °C.

4. Finite element analysis

Finite element analysis was carried out using LUSAS v14.5 finite element software. The cylindrical structure was idealised using two-dimensional axisymmetric models comprising quadrilateral solid field and continuum elements. A maximum element size of 0.3 m was established following a mesh convergence study. The thermo-mechanical transient analysis involves a coupled procedure and is time-stepped according to pre-defined intervals. The thermal analysis, which runs first, is governed by the quasi-harmonic transient heat conduction equation. The resulting temperature distribution for a given time step is subsequently fed to the structural analysis for the calculation of displacements and consequent stresses and strains. The material properties used throughout the finite element study are given in Table 1.

Material property	Value
Modulus of elasticity	33.0 GPa
Poisson's ratio	0.2
Coefficient of thermal expansion	$10 \times 10^{-6} / \text{K}$
Thermal conductivity	1.5 W/mK
Specific heat	900 J/kgK
Convective heat transfer coefficient	720,000 J/m ² dayK

Table 1: Material properties used in finite element study

Figure 4 shows the mesh discretisation for the axisymmetric models and displays a deformed contour of hoop forces obtained from the structural analysis.



Figure 4: Finite element model showing a deformed contour of hoop forces

The modelling procedure was verified with existing results in the literature from Priestley (1976), Ghali and Elliott (1992) and Vitharana and Priestley (1999). Good agreement for hoop forces and vertical bending moments was observed in each instance. Models with pinned bases and free at the top were used for the validation, with an ambient temperature gradient of 30 °C. Figure 5 shows the finite element model and associated internal forces for the comparison with Priestley (1976) for a tank of dimensions $R = 15.1$ m, $H = 7.2$ m and $t = 0.2$ m.

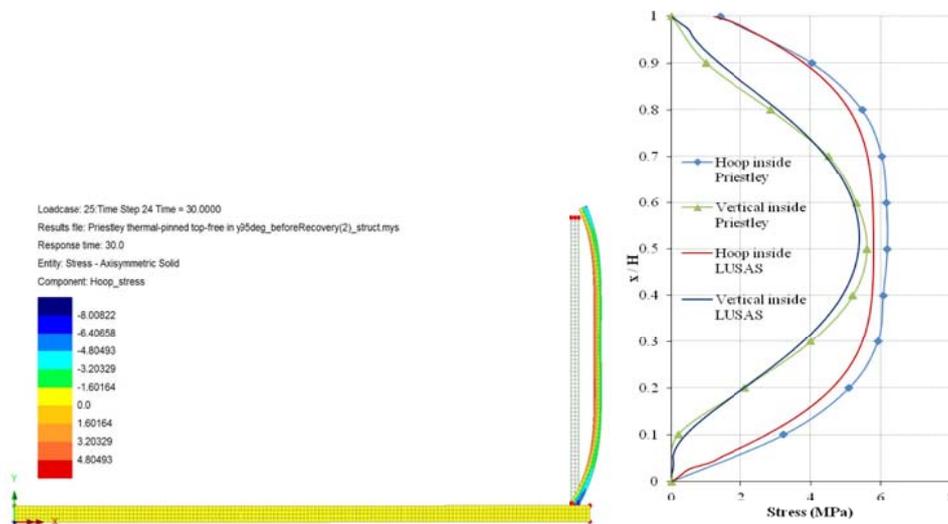


Figure 5: Deformed contour of stresses and associated internal force distribution for comparison with Priestley (1976)

5. Design implications

5.1. Limiting compressive stress

For a tank wall restrained radially at its ends, the resulting thermally induced hoop force is compressive over the entire height of wall, with a maximum occurring at the ends. Table 2 shows compressive stress limits from standards and guidance including EN 1992-1-1 (2004), BS 8007 (1987), PCI (1987) and NZS 3106 (1986). The limits are expressed in terms of the concrete cylinder strength, f_{ck} . The limit stipulated by BS 8007 (1987) is given in terms of the cube strength, f_{cu} , but an approximate conversion is made here.

Design standard/guidance	Limiting compressive stress at service
EN 1992-1-1	$0.45 f_{ck}$
BS 8007	$0.41 f_{ck}$ ($0.33 f_{cu}$)
PCI	$0.45 f_{ck}$
NZS 3106	$0.40 f_{ck}$

Table 2: Limiting compressive stresses in service

Figure 6 gives hoop compressive stresses for various temperatures for a tank with dimensions $D = 30.4$ m, $H = 7.0$ m and $t = 0.2$ m. A concrete strength class of C40/50 and a modulus of elasticity of 33 GPa is assumed. Since the temperature distribution for the current study involves predominantly a constant temperature across the concrete section, for simplicity, the small differential may be ignored.

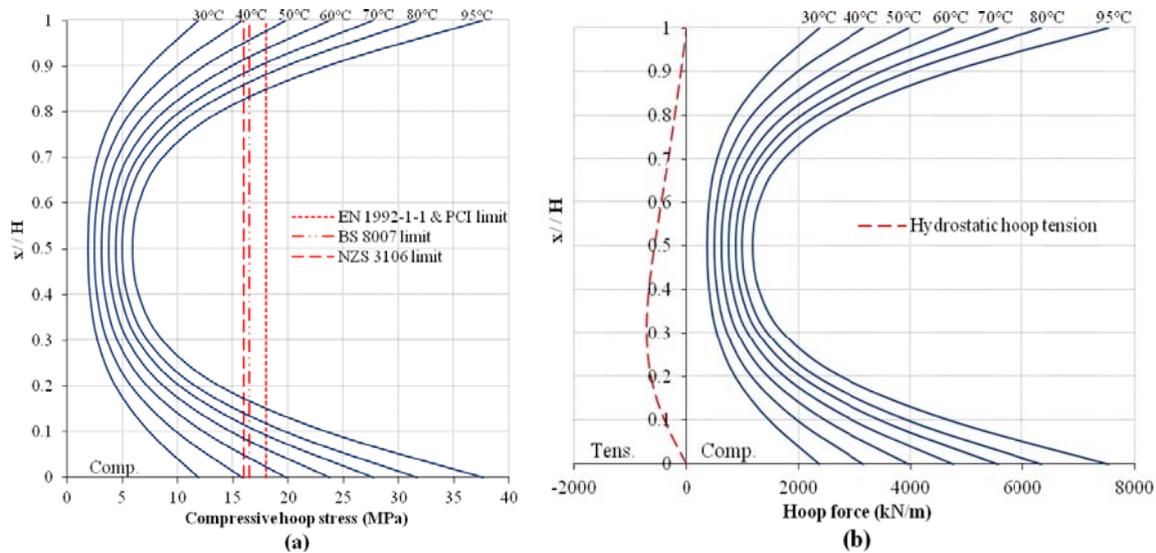


Figure 6: Compressive hoop forces and hoop stresses for various temperatures

Figure 6(a) shows that an average temperature of approximately 50 °C produces localised hoop compressive stresses at the wall ends that exceed each of the limits given in Table 2.

5.2. Circumferential post-tensioning

As the thermally induced hoop forces over the wall height are compressive, the circumferential post-tensioning requirements remain unchanged. The hoop tension arising from hydrostatic loading when the liquid is not heated is the critical loading condition that the circumferential post-tensioning is designed to cater for. Figure 6(b) includes the tensile force distribution resulting from hydrostatic loading derived using beam-on-elastic foundation analogy.

5.3. Vertical prestressing

The vertical bending moment distribution arising from hydrostatic and thermally induced loading is given in Figure 7. In order to establish an approximate limit on the vertical moments, a cracking moment was calculated from the following equation:

$$M_{cr} = (f_t + f_{max}) \left(\frac{I}{t/2} \right) \quad [1]$$

I is the second moment of area, t is the wall thickness, f_t is the concrete tensile strength and f_{max} is the concentric pre-compression stress required to eliminate tensile stresses whilst also satisfying maximum compressive stress limits at the extreme fibre. Again, as was the case previously, an average temperature across the concrete section of approximately 50 °C is prohibitive and produces vertical moments that exceed the cracking moment.

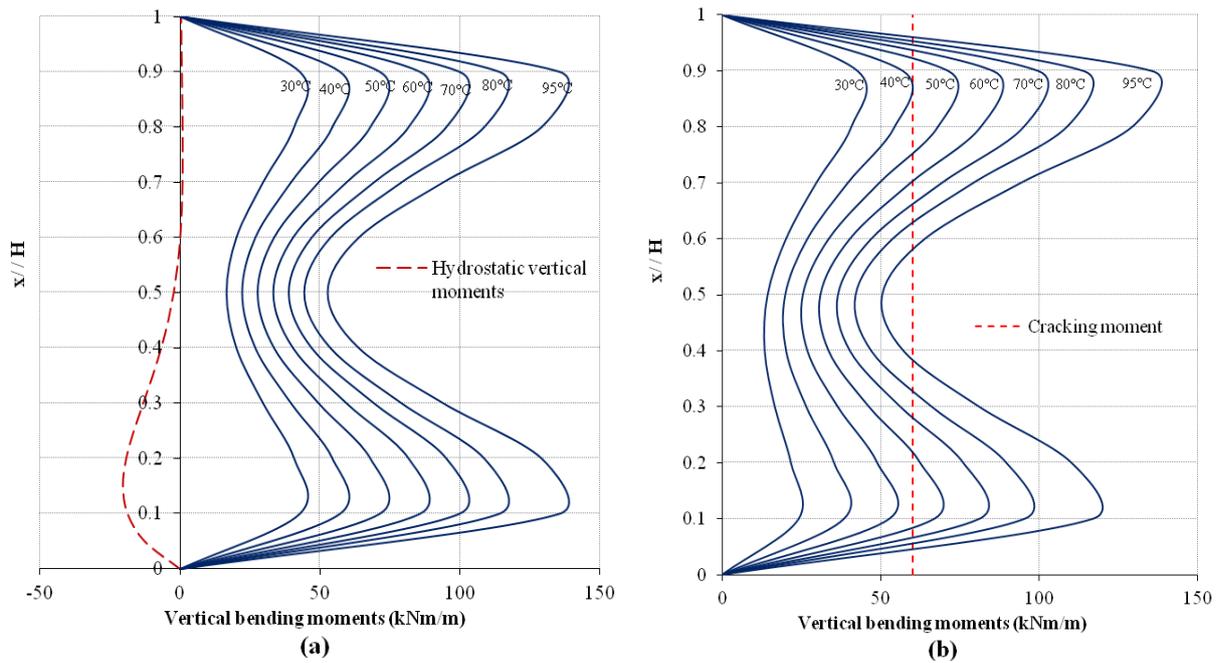


Figure 7: Vertical bending moments for various temperatures

5.4. Buckling

It has been established that thermal loading subjects restrained tank walls to significant combined axial compression and bending. Since prestressed concrete tanks are essentially shell structures, buckling stability should be addressed. For relatively stiff structures, linear eigenvalue buckling analysis is a technique that can be applied to approximate the maximum load that can be sustained prior to structural instability or collapse. The underlying assumptions of a linear eigenvalue buckling analysis are that the linear stiffness matrix remains unchanged prior to buckling and the stress stiffness matrix is a multiple of its initial value. Accordingly, provided the pre-buckling displacements have an insignificant influence on the structural response, the technique can be used effectively to predict the load at which a structure becomes unstable.

LUSAS v14.5 finite element software was used to carry out the linear eigenvalue buckling analysis. The three dimensional models comprised thick shell elements. The thermally induced compressive hoop forces and vertical bending moments, derived from the axisymmetric modelling, were simulated using a combination of internal stress-strain loading and externally applied radial pressure loading. The loading is also factored using partial safety factor of 1.55 for persistent thermal actions in accordance with EN 1990 (2002) and EN 1991-1-5 (2003). For plate or shell structures, it is prudent to include an initial geometric imperfection, as the buckling load is often sensitive to any deviation from the true geometry. Bradshaw (1963) made efforts to measure concrete cylindrical shells in the field and concluded that imperfections were observed to be as large as the shell thickness. Therefore, for the current study, an initial geometric imperfection of order of magnitude of the shell thickness was adopted.

The mode of buckling obtained from the finite element analysis is given in Figure 8, with the same mode observed for all tank sizes. The buckled shape displays the characteristic sinusoidal buckle waves consistent with similar studies in the literature.

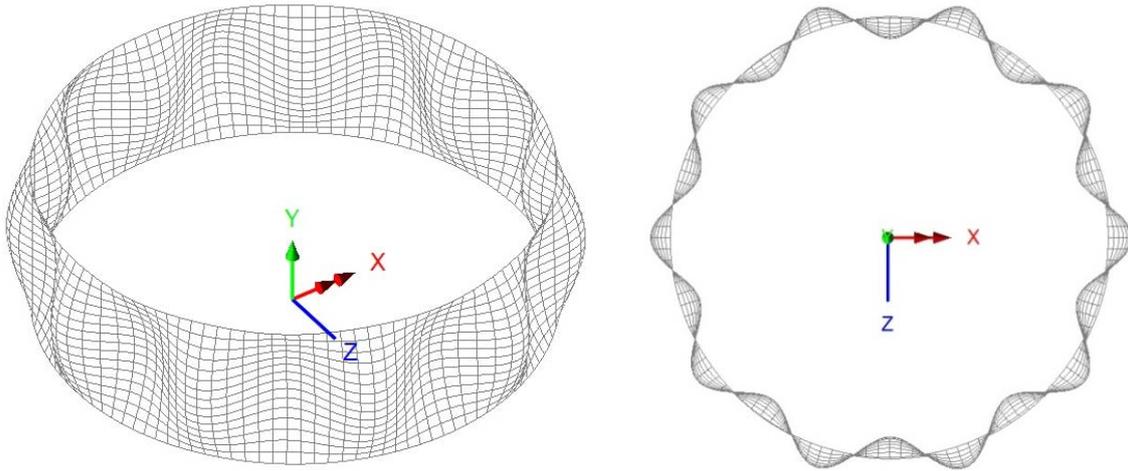


Figure 8: Buckled shape of pin-ended cylinder subject to thermally induced combined axial compression and bending

Figure 9 presents the eigenvalues extracted from the finite element study for various average temperatures and H^2/Dt ratios. The eigenvalues, λ , are ratios of the buckling load to the applied load. An eigenvalue equal to unity indicates that structural instability or buckling has occurred.

It is apparent from the resulting eigenvalues that for a pin-ended tank wall subject to thermal loading, the ultimate limit state of buckling is a necessary consideration, particularly for tanks with large H^2/Dt ratios.

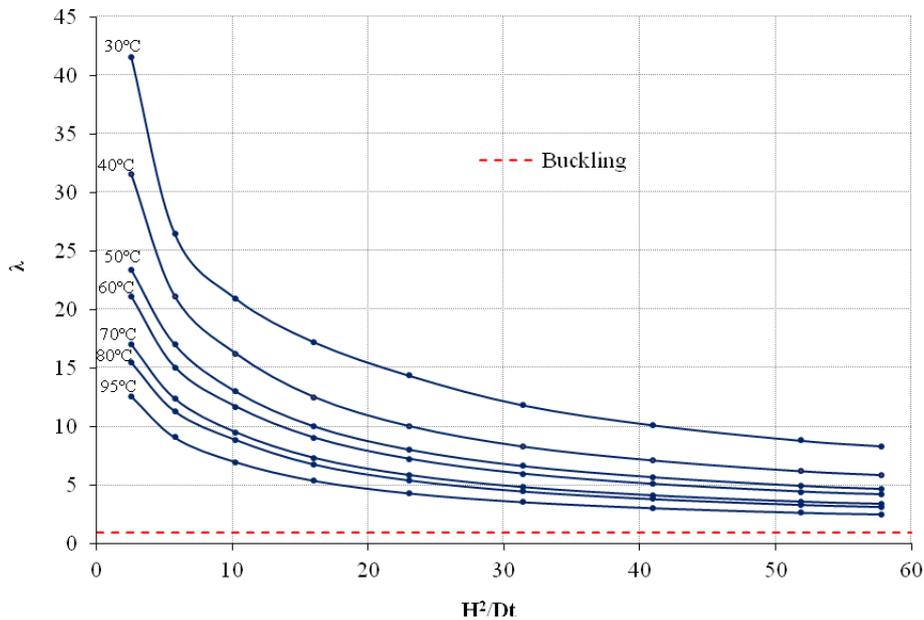


Figure 9: Resulting eigenvalues from finite element linear buckling analysis

5.5. Free-sliding condition

Theoretically, for a free-sliding condition, an average temperature across the concrete section does not induce any additional stresses. For a gradient experienced across the wall thickness, however, associated hoop and vertical bending stresses develop. Figure 10 is an example for a free-sliding wall subject to a temperature distribution resulting from the storage of heated liquids. The inside and outside temperatures are taken as 95 °C and 80°C respectively. This arrangement is slightly conservative as the use of external insulation would generally result in a temperature difference between the inside and outside faces less than 10 °C.

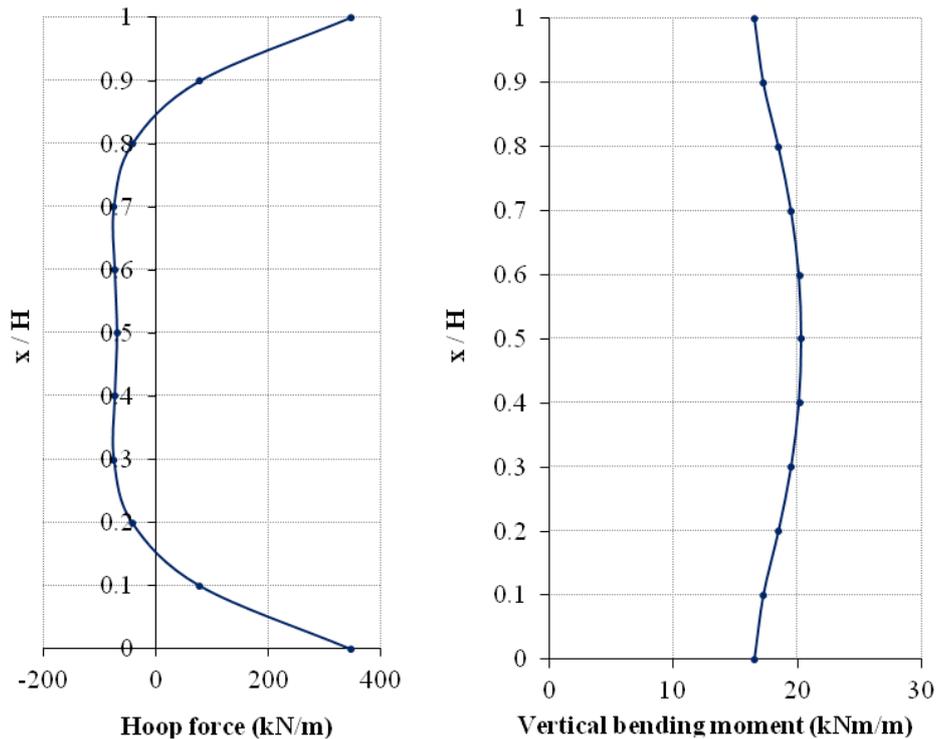


Figure 10: Internal forces resulting from thermal storage for a free-sliding wall

Comparing the results observed in Figure 9 with those in Figure 5 and Figure 6 it is evident that the magnitude and significance of the internal forces are far less for a free-sliding wall. A noteworthy observation is the hoop tension developed over the majority of the wall height which, although not excessive in magnitude, would need to be summed to the hydrostatic hoop tension when calculating circumferential post-tensioning requirements.

6. Conclusions and recommendations

For the most part, the temperature under consideration for the present study does not have a significant adverse effect on the material properties. The most important factors that require consideration are creep of the concrete and bond strength and stress relaxation for pretensioned and non-pretensioned reinforcement.

For a tank wall restrained radially at its ends, the internal forces resulting from the storage of heated liquids have been shown to be significant. As such, a temperature exceeding approximately 50 °C across the concrete section appears to be prohibitive based upon compressive stress limits and vertical prestressing constraints. Consequently, for a wall restrained radially, it is recommended internal insulation be provided to prevent temperatures from exceeding this. A linear eigenvalue buckling analysis has revealed that the ultimate limit state of buckling of pin-ended walls requires consideration, owing to the simultaneous axial compression and bending resulting from the thermal loading.

Where allowances are made for radial displacements in service, the 95 °C maximum temperature does not induce excessive stresses. Complications may arise, however, surrounding possible leakage at the joints. Accordingly, it is recommended that a heavy duty polymer liner be included internally, thereby eliminating concerns regarding liquid-tightness.

7. References

- AS 3735 (1991), "Concrete Structures for Retaining Liquids", Standards Australia, 40p.
- Bazant, Z. P., and Kaplan, M. F. (1996), "Concrete at high temperature: material properties and mathematical models", Longman Group Ltd., England.
- Bradshaw, R.R. (1963), "Some Aspects of Concrete Shell Buckling", *ACI Structural Journal*; 60(3), pp. 313-328.
- Brown, R.D. (1968), "Properties of concrete in Reactor Vessels", *Conference on Prestressed Concrete Pressure Vessels*, Westminster, pp. 131-151.
- BSI (British Standards Institution). (1987), "BS 8007: Code of practice for design of concrete structures for retaining aqueous liquids", BSI, London.
- Chiang, C., and Tsai, C. (2003), "Time-temperature analysis of bond strength of a rebar after fire exposure", *Cement and Concrete Research*, Vol. 33, pp. 1651-1654.
- Comites Euro-International du Béton (CEB). (1991), "Fire design of concrete structures – in accordance with CEB/FIP model code 90", *CEB Bulletin D' Information 208*, Lausanne, Switzerland.
- Fédération Internationale du Béton (FIB). (2010), "FIB Bulletin 55: Model Code 2010 – Volume 1 - Draft", Lausanne, Switzerland, 3 17p.
- Ghali, E. and Elliott, E. (1992), "Serviceability of Circular Prestressed Concrete Tanks", *ACI Structural Journal*; 89(3), pp. 345-355.
- Gross, H. (1975), "High-temperature creep of concrete", *Nuclear Engineering and Design*; 32(1), pp. 129-147.
- Haddad, R. J., Al-Saleh, R. J., and Al-Akhras, N. M. (2008), "Effect of elevated temperature on bond between steel reinforcement and fibre reinforced concrete", *Fire Safety Journal*, Vol. 43, pp. 334-343.
- Hannant, D.J. (1968), "Strain Behaviour of Concrete Up to 95°C under Compressive Stresses", *Conference on Prestressed Concrete Pressure Vessels*, Westminster, pp. 177-192.
- Harada, T., Takeda, J., Yamane, S., and Furumura, F. (1972), "Elasticity and Thermal Properties of Concrete Subjected to Elevated Temperatures", *American Concrete Institute*; SP-34, Vol. 1, pp. 377-406.
- Huang, Z. (2010), "Modelling the bond between concrete and reinforcing steel in a fire", *Engineering Structures*, Vol. 32, pp. 3660-3669.
- IS EN 1990. (2002), "Basis of structural design", NSAI, Dublin.
- IS EN 1991-1-5. (2003), "Actions on structures – Part 1-5: General actions – Thermal actions", NSAI, Dublin.
- IS EN 1992-1-1. (2004), "Design of concrete structures, Part 1: General rules and rules for buildings", NSAI, Dublin.
- Kagami, H. (1975), "Properties of Concrete Exposed to Sustained Elevated Temperatures", *3rd International Conference on Structural Mechanics in Reactor Technology*; London, Paper H1/5.
- Kagami, H., Okuno, T., and Yamane, S. (1975), "Properties of Concrete Exposed to Sustained Elevated Temperatures", *3rd International Conference on Structural Mechanics in Reactor Technology*; Paper H1/5, London, pp. 1-10.
- McDonald, J.E. (1978), "Creep of Concrete under Various Temperature, Moisture and Loading Conditions", *Douglas McHenry International Symposium on Concrete and Concrete Structures*, American Concrete Institute, Vol. 1, pp. 377-406.
- Nasser, K.W. and Neville, A.M. (1965), "Creep of Concrete at Elevated Temperatures", *Journal of the American Concrete Institute*: Vol. 62, pp. 1567-1579.
- NZS 3106 (1986), "Code of Practice for Concrete Structures for Retaining Liquid", Standards Association of New Zealand, 77p.
- PCI Committee on Precast Prestressed Concrete Storage Tanks. (1987), "Recommended Practice for Precast Prestressed Concrete Storage Tanks", *Journal of the Prestressed Concrete Institute*; 32(4), pp. 80-125.
- Priestley, M.J. (1976), "Ambient Thermal Stresses in Circular Prestressed Concrete Tanks", *ACI Structural Journal*; 73(10), pp. 553-560.
- Tadros, M. and Joo, Y. (1987), "Precast Concrete Storage Tanks", *Concrete International*; 9(11), pp. 40-43.
- Vitharana, N.D. and Priestley, M.J. (1999), "Significance of Temperature-Induced Loadings on Concrete Cylindrical Reservoir Walls", *ACI Structural Journal*; 96(5), pp. 737-749.

The Development of a Low Cost Instrument for the Measurement of Tidal Stream and Run of River Flows

T. Daly, C. Gibbons, D. O'Reilly

Department of Mechanical Engineering
and National Maritime College of Ireland (NMCI)
Cork Institute of Technology, Cork, Ireland

Abstract

After an examination of current methods and the instruments available for the measurements of flows in open channels and tidal flows, a need for an effective low cost device was recognized. An off the shelf, high end, ultra-sonic speed detector which had been developed for general marine use was sourced and modified for data collection. A technique for the deployment of the instrument had to be developed so it would face into the oncoming flow at all times and that the deployment device itself would have a minimal effect on the velocity of the water flowing around it. Issues such as powering the device over prolonged periods and recording flow velocities and different depths would also need to be addressed.

Keywords

Tidal energy, Tidal flows, Open channel, Flow measurement.

1. Introduction

As the tidal energy industry develops there is increasing need to be able to ascertain the tidal flow velocities in areas where proposed tidal energy developments are contemplated. Accurate flow velocities are important to enable developers to determine the amount of energy available in certain areas and also to properly quantify the stresses that the equipment will be subjected to. Inaccurate flow information has led to the unexpected destruction of equipment which was not designed to withstand the forces involved. This results in very large costs for the developer in recovering the equipment, redesigning, repairing and redeployment. If the forces present due to excessive flow velocities were known in advance, many of these problems could be avoided. In the field of environmental studies and monitoring, accurate tidal flow velocities are also of great importance.

Open channel and tidal flow rates are inherently hard to measure due to variations in channel widths and depths along the length of the channel and due to the varying bathymetry of the sea bed. Flow velocities also vary greatly with depth due to drag caused by channel bottoms and sides.

2. Earlier Studies

Earlier studies looked at the different methodologies and equipment for the measurement of flow velocities in both open channels and tidal areas. The following is a brief synopsis of these methods.

2.1. Weirs

A weir is a simple arrangement for measuring flow rates in open channels. Once the flow rate is established, the flow velocity can then be calculated. Weirs are simple and accurate and work by diverting the flow through a structure of known dimensions. This permits flow rates to be measured as a function of depth of flow through the structure. A simple weir consists of a bulkhead constructed from metal, concrete or timber with an opening cut into its top edge which is of known dimensions. The height of water in the weir is then measured at a given distance upstream of the weir and the flow velocity can therefore be calculated (Ackers, 1978).

2.2. Colour Velocity Measurements

Dyes have the advantage of being easily detected visually but are more accurate when used with fluorometer detection equipment. The dye is injected upstream and the water is sampled downstream. The degree to which the dye has been dissolved will depend on the amount of water that the dye has mixed with and is therefore a function of the flow rate. Multiple injection points and sampling at two stations downstream from the injection points produces very high accuracy (USDA, 2001).

2.3. Salt Velocity Measurements

Salt velocity measurements are conducted in much the same way as colour velocity measurements. In this case however, high concentrations of sodium chloride (NaCl) are injected into the flow to be measured. Higher concentrations of salt increased the electrical conductivity of the water. Pairs of electrode placed downstream record the conductivity of the water as it passes (USDA, 2001).

2.4. Anemometer and Propeller Current Meters

Anemometer and propeller current meters consist of a propeller or anemometer cups which rotate on a shaft when placed in a current and each revolution of the propeller produces a pulse or a number of pulses. The faster the propeller turns, the more pulses are produced. The speed at which these pulses are produced determines the speed of the flow the instrument is placed in. These instruments are directional and must face into the oncoming flow in order to work (USDA 2001).

2.5. Deflection Meters

A deflection meter consists of a shaped vane that is placed into the flowing water. The velocity of the flowing water exerts a force on this vane and a secondary device such as a load cell measures the deflection of the vane. The vanes can be shaped to match the flow section geometry to make them deflect the same amount for any given discharge regardless of the depth of flow in the flow section.

2.6. Pitot Tubes

Pitot tubes consist of a simple tube pointing directly into the fluid flow. The flow of fluid entering the tube results in a pressure build up within the tube. The tube has no outlet and therefore the fluid comes to rest within the tube. This pressure can then be measured is a function of the velocity of the fluid acting on the inlet of the tube.

2.7. Doppler Type Acoustic Meters

Doppler type acoustic meters consist of an ultrasonic transmitter and a receiver usually incorporated in a single head. Signals are transmitted at a known frequency and these signals reflect off suspended solids in the fluid. The receiver then picks up this reflection and the frequency shift of the received signal from that of the transmitted signal is related to the mean velocity of the fluid (Wanis and Hull, 2011).

2.8. Transit-Time Flow Meters

Transit time ultrasonic flow meters are based on the principle that the transit time of acoustics signals along a known path is altered by the fluid velocity. An ultrasonic signal sent upstream travels slower than a signal sent downstream. By transmitting an ultrasonic signal in both directions along a diagonal path, the average path velocity can be measured and the average flow velocity can be therefore determined (Baker, 2000).

2.9. Cross-Correlation Ultrasonic Meters

A cross correlation meter consists of two ultrasonic transmitters and receivers. The transmitters and receivers are arranged in two pairs with a small distance between them. The instrument works by transmitting an acoustic signal at high frequency which reflects back to the paired receiver from suspended solids in the water. The second transmitter also transmits an identical signal which is reflected back to its paired receiver. The profile of the signals received by both receivers is compared and when a signal at the second receiver is the same as a signal previously received by the first, then the same column of water has passed beneath both receivers. Because of the known distance between two receivers, it is possible to determine the time taken for the water column to travel from one point to another and therefore the flow velocity can be determined (www.shintech.com).

2.10. Measurement by Float

This is a method of flow velocity measurement as against an instrument for flow velocity measurement. A float is placed in river at a place where it is straight and uniform in cross section and grade with a minimum of surface waves. The floats are then allowed to travel between two measuring points and the time taken is recorded.

2.11. Hall Effect Paddle Wheel Measuring Device

This is an instrument which was earlier developed during this project. It consists of a Hall Effect transducer incorporated in a paddle wheel which turns when placed in moving stream of water. As the paddle wheel turns, it emits a pulse and the frequency of these pulses determines the speed of the paddle wheel and therefore the velocity of the stream of water. The frequency of the pulses is recorded and this information can be collected over a period of hours to weeks and can then be downloaded to a computer for analysis.

3. Reasons for Further Development

The need for a device that could record accurate flow velocity information over prolonged periods cost effectively is very desirable. Ideally this instrument should be capable of measuring flows without being fouled easily. It should be able to record flow velocities at predetermined depths or ideally at a range of predetermined depths as flow velocities can vary greatly with depth. It should have sufficient power to enable it to operate over long periods as tidal data is often required over long periods. Ideally it should record information on velocity, depth and direction. It should also be easily deployable and recoverable. Deployment in conjunction with submerged equipment such as tidal generators on a permanent or semi permanent basis would also be desirable.

All of the above flow velocity measurement methods are useful in certain applications but are unsuitable for others. Some are very difficult to implement in the field and do not give consistently accurate results for a variety of reasons. The equipment required in some cases is very expensive and requires specialist training to analyze the resultant data. This is especially the case when measuring tidal flows.

4. Approach Taken

The first task was to research and obtain a sensor that would be capable of measuring the flow velocity with a large degree of accuracy. A marine grade cross correlation ultra-sonic sensor with a resolution of 0.05m/s was obtained. This sensor is a solid state device which does not require any calibration and should therefore operate accurately with minimal maintenance. This sensor is a relatively heavy device and has associated electronics. The output from this sensor is a pulse which varies in frequency with flow. The sensor itself requires 185mA to operate. This poses a challenge in itself. The sensor is also directional which means that it must face into the flow at all times.

These are all drawbacks that must be overcome but the robustness and accuracy of the sensor make the viability of overcoming these problems worth investigating.

5. Design and Development

A method of mooring the device had to be developed that would allow it to move up and down and also to allow it to rotate so that it would face into the oncoming tide at all times. After a number of attempts at trying to moor an earlier deployment device so that it would maintain a forward facing and stable aspect which involved suspending the deployment device in different ways from a float, it was decided that an alternative method needed to be developed.

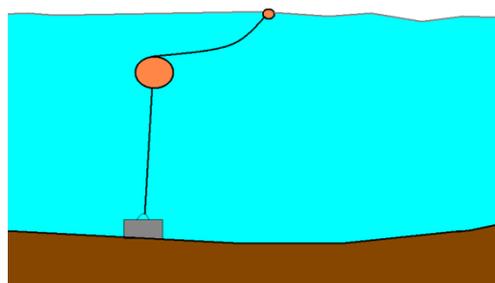


Figure 1: Mooring System

The system developed was a weight with a large buoy attached to it by means of a length of rope which would be shorter than the depth of water at low tide. This would ensure that the buoy would remain submerged at all times. This would in turn ensure that the rope between the weight and the buoy would at all times be taut. A smaller buoy would then be attached to the large buoy and allowed to float on the surface to warn mariners of the presence of something in the water.

The device would then have an attachment at the front which would incorporate a ring through which the taut rope would pass. This would allow the device to move up or down on the rope as required and also allow it to rotate into the oncoming stream of water. Initial tests with an earlier device moored in this fashion proved successful however further tests will be required. Due to the difficulties in testing equipment in tidal environments and the costs involved, much of the testing of the device and ancillary equipment will have to be carried out simultaneously at a later time and therefore only rudimentary testing at each stage of the development was carried.

An instrument had to be designed to deploy the sensor so that it could be submerged to take tidal flow velocity measurements at different depths. To do this, the device would need to house the sensor and the associated electronics. A battery capable of powering the sensor would also need to be housed in the device along with data logging equipment. The device would need to be streamlined and stable when submerged in a tidal stream.

The first design appeared to meet most of these requirements. It consisted of an oval shaped device with vertical hydroplane to aid stability. This design allowed for the inclusion of two ultra-sonic sensors as it was thought that an average of both measurements would provide a more accurate result. The body of the device was of sufficient size to house the electronics and a pair of 7Ah, 12V lead acid batteries. The flange used to join the upper and lower parts would act as a horizontal hydroplane, again increasing horizontal stability.

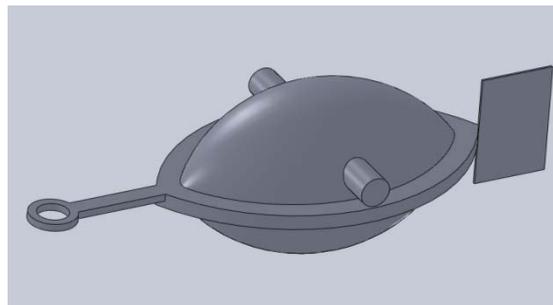


Figure 2: Initial Design

The design shown in Figure 2 incorporates the mooring attachment described earlier. This design and any subsequent designs would need to incorporate an adjustable buoyancy tank so that the depth at which flow velocity measurements would be taken could be either pre-set or programmable.

5.1. CFD Analysis

The ultra-sonic sensors chosen to take the measurements operate by transmitting an ultra-sonic signal and this signal reflects off suspended particles in the water. The reflection returns a profile dependent on the suspended particles in the water. The sensor contains two transmitter-receiver pairs positioned a known distance apart. When a reflected profile at the second receiver matches a profile already received at the first receiver, then the same column of water has passed both transmitter-receiver pairs and therefore the velocity at which it has travelled can be calculated. The sensors measure the velocity of the water in an area 77mm to 127mm from the face of the sensor. It is therefore very important that the device used to deploy the sensor does not affect the measurement. This could occur if the shape of the deployment device causes the fluid to speed up excessively as it passes around the device. Any shape that is placed in a flow of water will result in pressure changes on the surface of the object. Bernoulli's principle is that as the pressure increases the flow will decrease and where the pressure decreases the flow will increase.(Y. Nakayama 1999) This can be seen in Figure 3 where a simulation of a sphere of 80mm radius is placed in a flow of water at 2m/s. On the front face of the sphere the pressure is at its highest and the streamlines show that the velocity of the flow is lowest whereas around the center of the sphere where the pressure is lowest as indicated by the white area, the velocity of the flow is at its highest as indicated by the colour of the streamlines at this area. This must be considered when designing a deployment device for the ultra-sonic sensor.

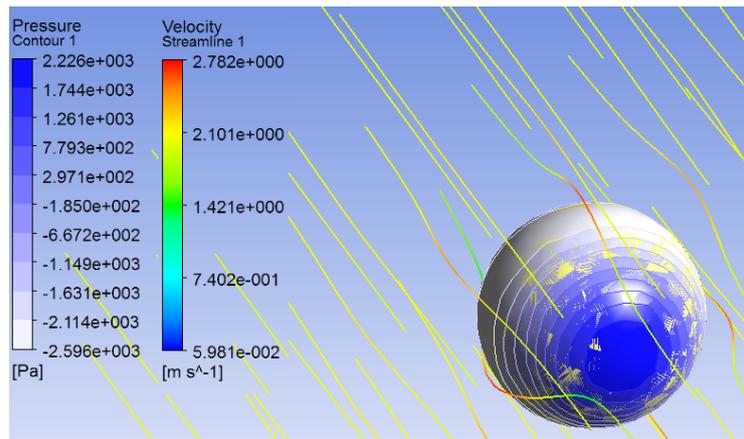


Figure 3: Analysis of Pressure v Velocity

While a disturbance in the flow velocity is inevitable when a body is placed in a flow, efforts must be made to minimize this disturbance so that the area where the velocity is measured by the sensors is not affected. CFD analysis was carried out on the first design to ascertain how it would affect the flow velocity in the critical measuring area. The flow was set at 2m/s as this is seen as a benchmark, minimum speed for tidal energy production at the moment. In post CFD analysis a plane was set up perpendicular to the faces of the sensors and different velocity thresholds were set on this plane. This would allow a visual result of where the velocity exceeded the pre-set threshold on the plane. Figure 4 shows the results of this analysis. The velocity threshold on the plane was set at 2.2m/s as this would allow for an error of 10% in velocity measurements.

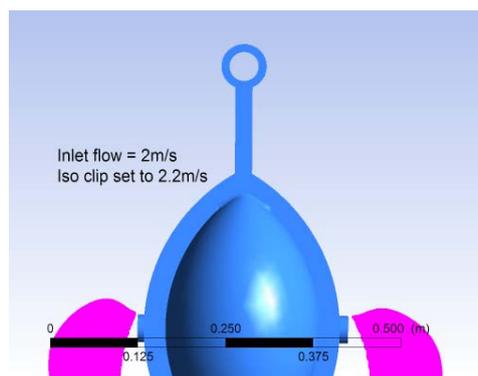


Figure 4: CFD Analysis of Initial Design

10% would be an unacceptable error however this Figure was chosen as a beginning point. It can be seen from Figure 4 that the area where the flow disturbance exceeded a velocity of 2.2m/s extends 125mm from the face of the sensor. This would imply that any measurement taken would be at least 10% higher than the actual flow as the sensor measures between 77mm and 127mm from the sensor face. This design will therefore not work as the shape of the device is creating too much of a disturbance in the flow.

This design was modified by the addition of baffles in areas that it was thought would result in the flow hitting the object head on being deflected away from the face of the sensors. In most cases this actually resulted in an increase in velocity across the faces of the sensors. One of these designs is shown in Figure 5.

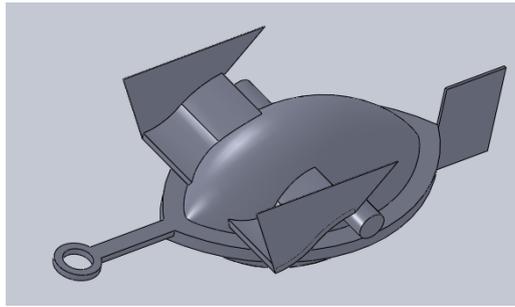


Figure 5: Initial Design with Baffles and Wings

The curved baffles on the top and bottom of the device were added in an effort to deflect the flow over and under the device. The wings added to the ends of both baffles were intended to maintain a laminar flow over the faces of the sensors. When this shape was analyzed using CFD it was found that the flow disturbance perpendicular to the sensor faces extended a distance of 125mm which was not an improvement on the original design. A number of variations on the same theme were tested but no significant advantage was observed. From these experiments it was established that the profile of the device facing into the oncoming flow was the characteristic which most influenced the flow disturbance around the shape. It was therefore decided to try to design a device which would have as small a profile as possible facing into the oncoming flow. The length of the device did not appear to have any discernible effect.

In order to design a device with a smaller profile some compromises with regard to the size of the batteries required to power the sensors would have to be made. This would have serious repercussions if the device was required to operate over long periods. This was a problem which would have to be considered at a later stage as the priority at this stage was to design a suitable shape for the device.

To try to determine the effect of an increase in flow on the disturbance that this would cause CFD analysis was carried out again on a simple shape. In this case a sphere was again used. The sphere was simulated in a flow of water at 2m/s and the results were analysed. A plane was set at the mid-point of the sphere and a velocity threshold of 110% of the flow was established. The distance from the sphere at which this velocity was exceeded was noted. The simulation was rerun at a flow of 3m/s and the threshold was increased to 110% of this flow. The distance from the sphere at which the velocity threshold was exceeded was found to be the same as with the lower flow.

This exercise was repeated for a number of different flows and it was found the distance from the sphere where the velocity exceeded 110% of the flow remained the same. There is a linear relationship between the increase in flow and the distance from the sphere where the disturbance occurs. This indicated that once a design was found that allowed the sensors to establish an accurate reading, any increase or decrease flow would not affect the percentage error of the measurement.

After extensive designs possibilities and CFD analysis the design shown in Figure 6 was arrived at. This device has a round body of 70mm diameter and extends to about 700mm in length. This size will allow the incorporation of the batteries to power the sensors and the electronics. The device also incorporates stabilizing hydroplanes attached to the rear of the body vertically and horizontally.

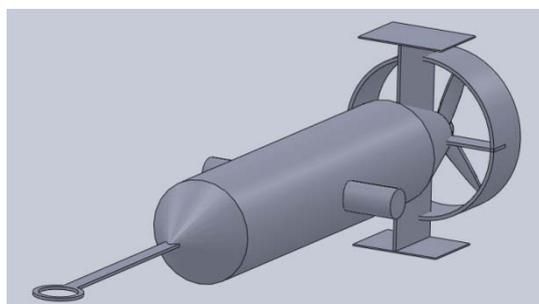


Figure 6: Final Design

This design kept the forward facing profile to a minimum while still allowing the device to accommodate a single 7Ah 12 lead acid battery. While this greatly reduced the power allowed for in previous designs, it was decided that there was not a need for the sensor to be operational continuously as any flow variation would be gradual and therefore it would

suffice for the sensors to operate for one minute in every fifteen. This would reduce the power requirements of the sensors by a factor of 15. If two sensors were to be used and they required around 190mA each, this would amount to 380mA per hour. However if the sensors only operated for a minute every fifteen minutes, this would reduce the power consumption to 25mA per hour. There would however be other equipment required on board such as a pressure sensor to determine the depth, an electronic compass to determine direction, timers to switch on and off the sensors, data logging equipment and a means of adjusting the buoyancy in the buoyancy tank if it was required to do this autonomously. The sensors alone could only operate for around twelve days and the device may be required to collect data over a much longer period. The addition of a propeller driven generator in the rear of the device would provide power for all of the onboard equipment and keep the battery recharged. Because the device would be anchored in the flow of water there should be ample power available to run a small generator and the power required to operate all of the equipment would be very small.

This design would also be easier to construct. It also allows for the device to be modular in that it could be made in a number of sections and assembled afterwards. This would allow a different section to house different components. The advantage of this is that if any water ingress occurred at a certain point, the resultant damage to the onboard equipment would be limited to the area where the leak occurred and the rest of the equipment would be safe.

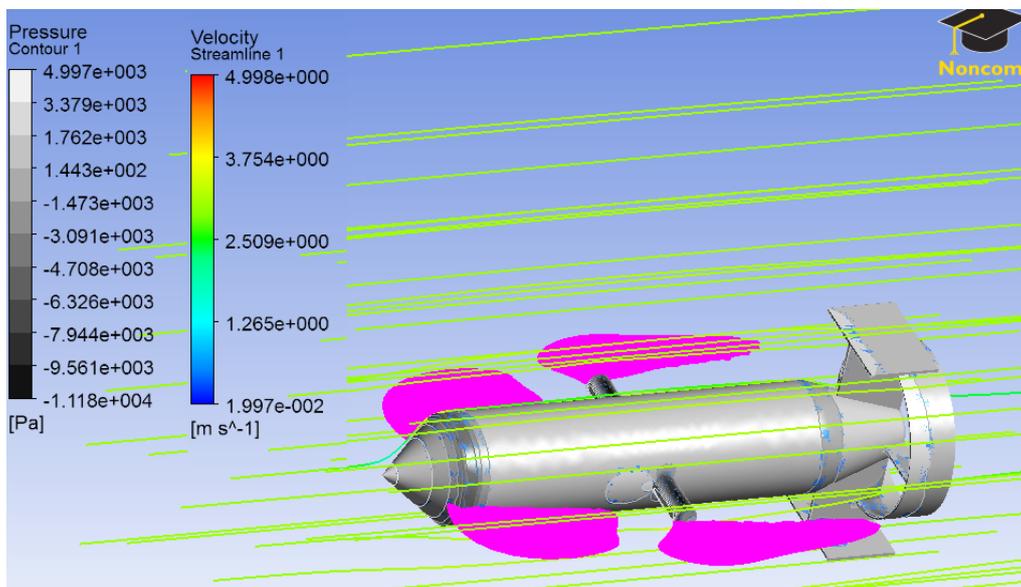


Figure 7: CFD Results of Final Design Simulation

CFD was carried out on this design on the same basis as earlier analysis and it was found that when a threshold of 103% of the flow was set on a plane perpendicular to the faces of the sensors, the distance at which the flow velocity increased beyond this threshold did not extend more than 70mm from the faces of the sensors. This analysis was carried out with varying flows and the results were conclusive with earlier test results in that distance did not change for a 103% threshold. This should allow for a measurement with an error of less than 3%. Figure 7 shows the results of CFD which was carried out on the device. The water flow was set at 3m/s and the velocity threshold was set to 3.12m/s which is 104% of the flow. The area where the velocity exceeded the threshold as a result of flow disturbance caused by the device itself only extended around 50mm from the faces of the sensors. The results also shows pressure contours over the entire surface of the device and it also shows velocity streamlines.

6. Model Verification

Due to the fact that there were no facilities availability to test the device in water it was decided to test a model of the device in a wind tunnel. The wind tunnel available had a test chamber of 30cm x 30cm x 1.5m. The Reynolds number for the device in an arbitrary flow of 2m/s was calculated but to achieve the same Reynolds number for a model of the device in a wind tunnel, the model would either need to be very large or the air flow would need to be very high. Because of the limitations of the wind tunnel available this was not possible so it was decided to construct a model of the device that would be suitable to the equipment available and to model a similar sized device with similar air velocities with CFD. If the results were comparable then it could be assumed that the CFD results for an actual sized model in water would be accurate.

A model was produced in a rapid prototyping machine. Only half of the model was produced as it is symmetrical and therefore there is only need to test one half. The model was produced in five sections due to the limitations of the size of model the prototyping machine could produce.



Figure 8: Models Produced in the Rapid Prototyping Machine

The five parts allowed the model to be tested in two configurations, one being 90° offset from the other. Only three parts were used for each configuration with the front section being common to both. This was done to reduce the time taken for the rapid prototyping machine to produce the models and to limit the cost involved. These models can be seen in Figure 7. The front section of the model on the right will be fitted to the model on the left when required.

This model will be tested in the wind tunnel. Pressure readings at different points on the model will be taken. A smoke generator will be employed to enable the visualization of the airflow over the model and cotton strings will be attached to the model to indicate the direction of the streamlines. This data will then be compared to the CFD results obtained with the same simulated conditions.

7. Conclusion and Future Plans

The sensor that has been identified as suitable for this device is a highly accurate, low cost sensor designed for use in the marine industry. The device itself is a simple tubular construction designed to hold the sensor electronics, the data logger and the batteries. A very small power generator incorporated into the tail of the device would be sufficient to keep the batteries charged and therefore allow the device to operate for long periods. The data logger used is also a low cost instrument capable of gathering frequency data for long periods, depending on the sampling rate. The proposed mooring system is very simple and is easily recoverable for reuse. The addition of on-board equipment such as pressure sensors and an electronic compass would add to the complexity and cost of the device but this equipment would add significant capability to the device.

The work done to date would suggest that a relatively cheap, accurate and sophisticated device could be manufactured. This device could do the work of much more expensive equipment and therefore allow areas to be surveyed for small scale tidal energy projects without making the cost of the surveying prohibitive and therefore putting the whole project in jeopardy.

It is planned to carry out wind tunnel testing shortly to try to verify the CFD modeling. If the CFD modeling compares closely with the wind tunnel testing, then it can be assumed that the full scale simulation of the model in water will also be accurate.

Some additional testing of the sensors will be carried out in a tow tank to verify that the data logging equipment captures the output from the sensors correctly and that the data is interpreted properly so as to give an accurate measurement.

The next part of the project is to look at the buoyancy requirements and to examine methods for an autonomous buoyancy system to allow the device to take measurements at different depths over time and a system of transmitting data to shore on a synchronised basis which would allow for up to date velocity readings being available without the need to recover the device to download data.

8. References

Ackers, P. W., Perkins, J. A., Harrison, A. J. M. (1978), Weirs and flumes for flow measurement, John Wiley & Sons Ltd.

Baker, R. (2000), Flow Measurement Handbook, Cambridge University Press.

USDA (2001), Water Measurement Handbook, U.S. Government Printing Office.

Wanis, P., Hull, M. (2011). New technology impacts on ADCP instruments. OCEANS 2011.

Nakayama , Y. (1999). Introduction to Fluid Mechanics, Arnold.

Geo-spatial Data Fusion for Anomaly Detection

J. Corkish², P. Walsh¹

¹Department of Computer Science
Cork Institute of Technology, Cork, Ireland
paul.walsh@cit.ie

²Research and Development Department
National Maritime College of Ireland, Cork, Ireland.
john.corkish@cit.ie

Abstract

This paper introduces an information fusion framework for maritime surveillance that integrates, assesses and visualises information in real time to provide an overview of potential threats to border security. The framework allows for integration of heterogeneous sensor information such as Automatic Identification Systems (AIS), Long Range Identification and Tracking (LRIT), RADAR, IR cameras sea bed terrain and historical records. The system utilises robust machine learning algorithms for real-time information classification of data into anomalous/non-anomalous behaviour. The logical machine learning architecture lends itself to parallel processing and the framework enables this through the use of multi-processing cloud based architecture that can scale up to meet the demands of real-time maritime scenario assessment.

Keywords

Data mining, anomaly detection, unsupervised learning, statistical learning, cluster based learning, machine learning

1. Introduction

The globe consists of approximately 70.78% water, covering 139.4 million square miles. Today, there are more than 135,000 commercial vessels, thousands of navy vessels, and countless small craft using global waters (Munoz, 2011). It is estimated that 90% of global trade and 50% of the world's oil are transported by the sea. This also includes 90% of the European Union's external trade and 40% of its internal trade. Europeans seaports and waterways account for three hundred and fifty million passengers and approximately 3.5 billion tons of cargo per year. Major chokepoints are the English Channel, the Danish Straits and the Straits of Gibraltar (NATO, 2012).

European borders consist of 8,826km land borders and 42,672km of external sea borders. In 2011 the European Commission established the European Border Surveillance System, (EUROSUR). The primary function of EUROSUR is to secure related maritime borders against illegal migration, illegal trafficking and other illicit maritime activities. The European FP7 project PERSEUS follows in the same context as EUROSUR. PERSEUS is a demonstrator project that unites 29 partners from 12 European countries. The objectivity of the project is to create a System of Systems (SoS). The aim of the project is to integrate various existing and future national maritime surveillance technologies. Maritime border and port protection activities are currently focused on building security capabilities required to protect against threats such as natural disasters, human rights violations, terrorism and crime (CORDIS, 2012). European member states envision linking existing national and regional control centres in order to increase border security and intelligent surveillance. The development of technological capabilities and security systems integration includes service processes and data sources, such law enforcement, civil defence and medical information infrastructures.

Maritime heterogeneous multi-sensor networks yield large quantities of varying data. National and regional control centres are responsible for monitoring multi-sensor data extracted from ground nodes, aerial reconnaissance and sea based platforms. Advances in technologies, high performance computing and data storage have consequently led to immense data sets. The ability to analyse such data sets represents a major challenge. The multidisciplinary field of scientific data mining is a mechanism that has exploratory benefits when analysing varied data sets. As data sets increase the need to identify anomalous or outlier data is imperative. Historically, data outliers were used to detect network intrusions and credit card fraud. (Chandola et al. 2009) states that an anomaly can be described as a deviation from the normality. Non conforming patterns can also be referred to as anomalies, outliers, discondant observations, exceptions, or aberrations. However, of all of these, anomalies and outliers are the two terms most commonly used in the context of anomaly detection. The ability to extract, process and analyse significantly large datasets is becoming increasingly important (Chandola et al. 2009).

Maritime Surveillance systems process and analyse vast amounts of heterogeneous sensor data. The dissemination of such extremely large multidimensional datasets into easily accessible information is pivotal to anomaly detection. Technologies may include radar, Automatic Identification Systems (AIS), Satellite Automatic Identification System (S-AIS), Long range identification and tracking (LRIT), vessel monitoring systems (VMS) and Infra-Red cameras. Underwater acoustic and electromagnetic networks also provide potentially undetectable observations. Additionally the introduction of Autonomous Underwater Vehicles (AUV's) and Remote Operated Vehicles (ROV's) provide detailed data that is both accurate and significant. Such technologies enhance the ability to establish a comprehensive maritime situation.

(Compieta et al. 2007), indicated that the ability to produce extremely high volumes of data has far exceeded exploratory and analytical capabilities. The integration and dissemination of information from varying sources is complex and the availability of quantum amounts of data represents an arduous challenge. One approach in overcoming this scenario is to create a robust and efficient automated “*priority data*” processing mechanism.

Threats come from both cooperative and non cooperative vessels. Cooperative: the acquisition of vessel identification messages (AIS) are sent either automatically or voluntary. AIS data is collected at national control centres/costal centres. Messages at national costal centres are also available from other platforms (patrol aircraft, patrol vessels and space borne sensors.). The correlation of such data sets enables the tracking of most of the cooperative vessels. Non-cooperative: The detection of fast and small boats, and other larger non cooperative targets requires a different approach. This may require the application of suitable algorithms for extracting anomalous vessel behaviour. It may also include the deployment of instrumented sensor platforms within ports and on the open sea. European projects are highly focused on improving detection capabilities by integrating systems capabilities, developing new data fusions algorithms and detection methodologies.

2. Background Information

Maritime surveillance and anomaly detection methods play an important role in the management and protection of sea borders and ports. Understanding the comprehensive requirements necessary for effective maritime surveillance can be difficult. Le Roux et al (2007) propose the concept the Joint Director of Laboratories (JDL) (Hall and Linas, 1997) data fusion model as a means of extracting system requirements for effective maritime surveillance. The revised (Steinberg, 2005) activity recognition framework fuses three components. Capability, intent and opportunity in order to predict and estimate the likelihood of an action. Based upon work conducted by (Steinberg 2007) the framework recognises any intentional behaviour instead of just threats or intentional threats. Maritime Domain Awareness projects have generally focused on methodologies that detect and track both cooperative and non-cooperative vessels. Often, such projects fail to include the importance of the human factor as a critical “feedback mechanism”. The human operator plays an important and yet mainly understated role in anomaly detection within the maritime security domain. (Endsley 1995) provides an indept study of situation awareness (SA), human cognetivity, and the importance of dynamic human decision making. Visual analytics provides valuable assistance to the analyst while assessing geo-spatial anomalies, as detailed by (Wilkinson et al. 2006). (Rhodes et al. 2007), discusses the effectiveness of combining human cognitive abilities with visual analytics technologies. Further discussion on the importance of the human interactive element can be found in (Song et al. 2007). The human feedback mechanism is an integral part of the machine learning process as indicated by (Riveiro et al. 2010). Due to its complexity, maritime anomaly detection methods tend to raise a high number of false alarms (in a real world context). Anomaly detection models that include the human expert knowledge as yet have not been fully explored. (Riveiro et al. 2008) indicated that anomaly detection systems present two key challenges: (i) they need to provide adequate user support and (ii) they need to involve the end user in the underlying detection processes. (Kimani et al. 2004), provides additional supportive knowledge in relation to the visual data mining environment. The (Rivera et al. 2009) paper extents this concept and concludes and by explaining that future development of anomaly detection systems should support and properly integrate the human element.

Anomaly detection in network security, traffic systems, credit card fraud and human activity monitoring is well documented in literature. In relation to Maritime Domain Awareness, there are fewer literature treatments and less real world practical applications. However, one such system is the Track Assessment and Anomaly Detection - Maritime (TAANDEM). The system is a track processing system used to identify motion that is inconsistent with previous behaviour patterns. In the context of Maritime Domain Awareness, the term anomaly can be clarified as events that required detection and identification. DARPA also worked on a system called “PANDA”. Certain types of data can be used in the creation a priority data-set. The dataset may include data ranging from AIS transmissions. Other forms of data may include a listing of crew members, the history of crew members, ship registration and ship's captain details. Data may also include shipping speed and port of origin, routes and port destination. (Laere and Nilsson, 2009), provide a further evaluation of maritime based anomalies systems. Suspicious vessel behaviour (alert/warning and incident occurred) is detected when a given combination of conditions (or rules) on track parameters and correlated information/intelligence is verified. Anomaly detection is an intense area of research with a diverse range of application.

Many anomaly detection methods focus on specific domains, conversely others are more universal in application. The (Chandola et al. 2009), *anomaly detection - a survey* provides a well-structured and comprehensive overview of research and methodologies. The importance of linking the human analytical reasoning with visualization and interaction techniques; is widely known as a powerful medium in relation to data mining (Riveiro et al. 2008). Linking domain expertise and powerful data mining engines enhances the analyst's cognitive ability. Thus increasing the ability to observe and deduce particular patterns (Kerren et al. 2007). The importance of the human operator and similar concepts are also explored by (Hunter, 2009).

2.1. Anomaly Algorithms and Approaches

Anomalous events tend to occur relatively infrequent. however, when an event occurs the consequences generally translate into negative impacts. The following illustrates anomalies in a simple two-dimensional data set (Chandola et al. 2009).

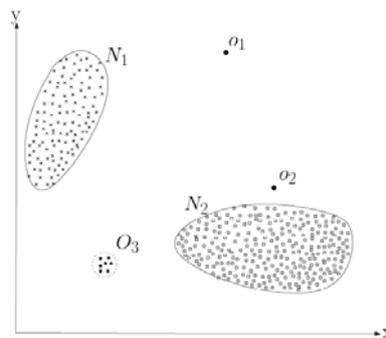


Figure 1 : An example of anomalies in a two-dimensional data set. (Hodge and Austin, 2004).

The data has two sectors N_1 and N_2 , notice that most observations lie in these two sectors. Points indicating distance from these two sectors are O_1 and O_2 and points in the sector O_3 are anomalies. Riveiro, et al, (2009), define anomalous as “not conforming to what might be expected because of the class or type to which it belongs, or the laws that govern its existence in a given situation or context. This definition is most applicable to the maritime domain. Vessel anomaly detection research area is of great importance within the maritime research community, (Auslander et al. 2011) (Riveiro, et al. 2009) (Laxhammar, 2008). There are many approaches and methodologies. One approach is autonomous detection and alerts systems as specified by (Riveiro and Falkman, 2010). The paper suggests the use of algorithms for the detection of event anomalies. These outliers can be classified in to three main groups: (i) Data driven-based, (ii) Signature/rule-based, (iii) Hybrid-based. It is important to note that the automation of learning and detecting vessel anomalies and event behaviours should also support the human decision making process. (Endsley, 1995), discusses and indicates that activity recognition and behaviour analysis assist future prediction. Cognitively inspired algorithms, outlined by (Rhodes, Bomberger and Zandipour, 2007) described motion pattern learning, analysis and prediction. Tests conducted yielded positive results in a number of real-life scenarios. Additionally, event-level normalcy learning and anomaly detection algorithms were demonstrated on a proto-type port and costal monitoring system developed and tested by (Rhodes et al. 2007). Furthermore; a rule-based maritime situation assessment ontology defined by (Edlund et al. 2006) used basic spatial and kinematical relations between objects to infer varying scenarios. The ability to detect anomalous behaviours requires a degree of predictability. Learning vessel behaviour patterns is context-specific and is dependent on elements such as vessel type, tides, weather conditions, location and speed.

2.2. Geomatic Spatial Zones

The study of costal zones and related sea areas can be defined as geospatial information analysis. A geographic information system (GIS) (Fotheringham and Wilson, 2008) is designed to capture, store, analyse and manipulate geographically referenced data. Geo-temporal Analysis visualizes maritime chat data in the X and Y axis and time in the Z axis in order to achieve geo-temporal patterns. Example of this includes GeoTime and Google fusion tables. Mapping of a coastal zones and sea areas is demonstrated by (Rhodes et al. 2007) whereby the port of Miami is subdivided into grid sections. A notable point was the indication that open sea areas leading to the port had not been addressed. The grid sections are used to provide co-ordinates reference points within a dataset. Similarly (Rhodes, 2007) used the neural associative incremental learning algorithm (NAIL) to fully automate a learning method based on neural network formalism. The model performance was tested against the SSE association rule mining algorithm proposed by (Kubica

et al. 2003). The Rhodes (2007) model used an error score representing the distance between a learned and the known network, calculated as follows;

$$SSE = SumA * SumB(w_{AB}^{true} - w_{AB}^{learned})^2$$

In the (Riveiro and Falkman, 2009) a hybrid model was developed in order to detect anomalous behaviour. The model used a combination of a data-driven method with a knowledge-based detection approach. This data driven method utilises two important variants; Self Organising Maps (SOMs) and Gaussian Mixture Models (GMMs) as discussed by (Riveiro et al. 2007). The data driven anomaly detection approach can be classified into statistical (parametric and non-parametric) machine learning. Bayesian networks, neural networks and clustering techniques are examples. The hybrid approach is a statistical parametric that combines SOMs and GMM, where the parameters of the Gaussian distributions (mean and covariance matrices) can be estimated from the available training data using SOM. Riveiro and Falkman (2009) define the normal behavioral model as a probability density function synthesized from a weighted sum of individual multivariate Gaussian distributions:

$$P(x_1, \dots, x_n) = \sum_{i=1}^D \sum_{j=1}^D P_{ij} P^{ij}(x_1, \dots, x_n)$$

The distribution P_{ij} , in this case correspond to the model vectors that were the output from the SOM. Each model vector is characterised by a n-dimensional Gaussian probability density function. The mean of each individual probability density function is given by the final weights for the model vector, while the variance is given by the dispersion of training data around the model vector.

3. A Taxonomy of Cluster based learning

The concept of clustering can be defined as the classification of data sets (patterns, observations or feature vectors) into groups. Data analysis procedures can be separated into two parts, confirmatory or exploratory. This is based on the model that is most appropriate for the given data set. Clustering can also be defined as “unsupervised classification” and discriminant analysis “supervised classification”. Data clustering is a useful exploratory tool for grouping, machine learning situations, data mining, image segmentation, decision-making and pattern classification (Jain et al. 1999). Its is a particularly appropriate model for the exploration of inter-relationships among data points in a multi-dimensional space.

Clustering is a technique used to group data sets that have similar instances. Clustering in general is an unsupervised method. However, semi-supervised and supervised are also used. There are several clustering based anomaly detection approaches. These can be grouped into three distinct sets. The first clustering category depends on the following assumption.

Assumption: normal data instances belong to a cluster in the data, while anomalies do not belong to any cluster. There are several clustering algorithms, (Chandola et al. 2009) indicate the following, DBSCAN (Ester et al. 1996), ROCK (Guha et al. 2000), and SNN clustering (Ertoz et al. 2003). These algorithms apply a clustering-algorithm to the data set and asserts that any data instance not belonging to the cluster as an anomaly. Another set of clustering algorithms remove the clusters from the data set, leaving only residue data. This data is classified as anomalous. (Chandola et al. 2009) papers indicates that these methods are supported by (Yu et al. 2002) with the find out algorithm, which in turn is an extension of the WaveCluster algorithm and the (Sheikholeslami et al. 1998) technique. A major disadvantage associated with these algorithms is the fact that they are not developed to detect anomalies but rather to detect clusters. The second cluster category relies on the following assumption.

Assumption: cluster anomalies are situated far from their closest cluster centroid, while normal data instances are situated close to their cluster centroid. Techniques that incorporate this assumption involve two steps. Firstly, a clustering algorithm is used to cluster data. Secondly, an anomaly is calculated by the from the data instance to its nearest cluster centroid, this distance is classified as its anomaly score. Using this two step method and different clustering algorithms several approaches have been proposed (Smith et al. 2002) revised *Self-Organizing Maps* (SOM), K-means clustering, and *Expectation Maximization* (EM). (Chandola et al. 2009).

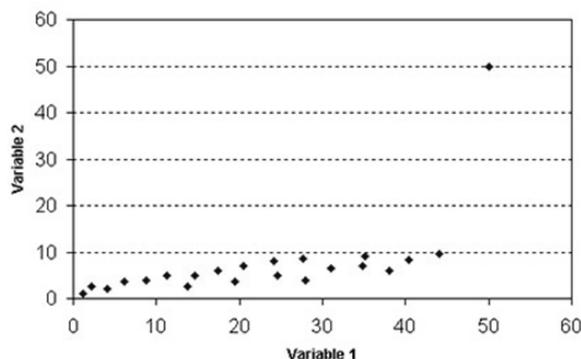


Figure 2: An example of a cluster data set with one defined anomaly. (Pang-Ning et al. 2005)

3.1. Supervised and Unsupervised Learning

In supervised learning, the learning algorithm uses labeled training examples from all classes in order to generate a classification function. An example of the functionality of a supervised learning can be stated as follows. Given a set of training data in the form $\{(x_1, y_1), \dots, (x_N, y_N)\}$, the learning algorithm strives for the function $g : X \rightarrow Y$, X is the input space and Y is the output space. The function g represents an portion of some space of possible functions G , this can also be referred to as the “hypothesis space”. Conveniently g can also be represented as the scoring function $f : X * Y \rightarrow \mathbf{R}$ so that g is defined as returning the y value producing the highest score: $g(x) = \arg \max_y f(x, y)$, letting f denote the space scoring function. The most popular and simplest unsupervised partitioning clustering strategy is founded on the square-error criterion.

(Jain et al. 2000) provide an overview of the functionality of this partitioning clustering strategy. The general objective is to attain a partition for a fixed number of clusters which will minimize the square-error. Suppose that the given set of n patterns in d dimensions has been partitioned into K clusters $\{C_1, C_2, \dots, C_k\}$ such that C_k has n_k patterns and each pattern is in exactly one cluster, so that

$$\sum_{k=1}^K n_k = n.$$

The mean vector, or center of cluster C_k is defined as the centroid of the cluster or,

$$m^{(k)} = \left(\frac{1}{n_k}\right) \sum_{i=1}^{n_k} x_i^{(k)},$$

Where $x_i^{(k)}$ is the i th pattern belonging to cluster C_k . The square-error for cluster C_k is the sum of the squared Euclidean distances between each pattern in C_k and its cluster center $m^{(k)}$. This square-error is also called the within-cluster variation.

$$e_k^2 = \sum_{i=1}^{n_k} (x_i^{(k)} - m^{(k)})^T (x_i^{(k)} - m^{(k)}).$$

The square-error for the entire clustering containing K clusters is the sum of the within-cluster variations:

$$E_K^2 = \sum_{k=1}^K e_k^2$$

The objective of a square-error clustering method is to find a partition containing K clusters that minimizes E_K^2 for a fixed K . The resulting partition has also been referred to as the minimum variance partition. An algorithm that applies the above partitioning strategy is the K-means algorithm. The K-means algorithm was discovered in the scientific field by (Steinhaus, 1956). Even though K-means was proposed over 50 years ago, it is still the most commonly used clustering algorithm to date. Its popularity of its success is due to the fact that it is easy to implement, it is simple, efficient, and has proven empirical success. The following is a summary of the K-means algorithm. Let $X = \{X_i\}, i = 1, \dots, n$ be the set of n d-dimensional points to be clustered into a set of K clusters, $C = \{c_k, k = 1, \dots, K\}$. The K-means algorithm finds a partition such that the squared error between the empirical means of a cluster and the points in the cluster is minimized. Let μ_k be the mean of cluster c_k . The squared error between μ_k and the cluster c_k is defined as;

$$J(c_k) = \sum_{x_i \in c_k} \|x_i - \mu_k\|^2$$

The goal of the K-means algorithm is to minimize the sum of the squared error over all K clusters.

$$J(C) = \sum_{k=1}^K \sum_{x_i \in c_k} \|x_i - \mu_k\|^2$$

The algorithm starts with an initial partition with K clusters and allocates patterns into clusters thus reducing the squared error. The squared error always decreases with an increase in the number of clusters K , (with K (with $J(C) = 0$ when $K = n$), this can be minimized only for a fixed number of clusters. (Jain and Dubes, 1988) state that there are five main steps to the K-means algorithm, there are;

1. Select an initial partition with K clusters; repeat steps 2 and 5 until the cluster membership stabilizes.
2. Generate a new partition by assigning each pattern to its closest cluster center.
3. Compute new cluster centers.
4. Repeat steps 2 and 3 until an optimum value of the criterion function is found.
5. Adjust the number of clusters by merging and splitting existing clusters or by removing small, or outlier, clusters.

The parameters of K-means requires three “user” specified parameters: the number of K clusters, cluster initialization, and the distance metric.

The most commonly used distance metric is the *Euclidean distance*, stated as;

$$d_2(x_i, x_j) = \left(\sum_{k=1}^d (x_{i,k} - x_{j,k})^2 \right)^{\frac{1}{2}} = \|x_i - x_j\|_2$$

(Jain, Murty, & Flynn, 1999) indicate that the Euclidean distance works very well in two and three dimensional space, particularly when used to calculate the distance between objects. The following figure illustrates the K-means algorithm on a two dimensional data set.

4. Related works

In a broad term data mining is the discovery of knowledge in databases (KDD) and is an iterative process of two parts: (i) data preparation and cleaning, (ii) hypothesis generation (data mining is usually based in this phase) and (iii) interpretation and analysis. Data mining techniques have been applied to intrusion detection systems (IDS) for quite some time (Anderson, 1972). Utilising data mining methods in the maritime situational domain includes concepts such as hidden Markov model (HMM), statistical and probabilistic techniques, (Carvalho et al. 2011) and artificial neural networks (Carpenter et al. 1992). Such methods have also been used in the detection of anomalous vessel patterns. (Bomberger et al. 2006), (Auslander et al. 2011) (Riveiro et al. 2009) (Rhodes et al. 2007).

The association rule (AR) learning/mining method is very effective in discovering variables in large data sets (Riveiro & Falkman, 2010) express that; an association rule is an implication of the form $A \rightarrow B$ where A is a set of *antecedent* items and B is the *consequent* item. Two commons methods of visualisation are: the two dimensional matrix (table based method) and the directed graph.

As shown in the following figure 1:

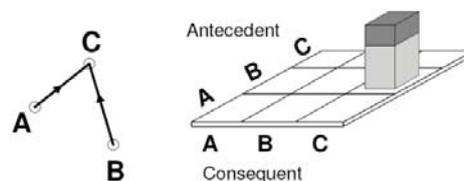


Figure 3: Basic representation of association rules: directed graph (left) and two dimensional matrix (right).

In the directed graph representation, the nodes depict the item and the edges represent the associations. Two association rules are shown: $A \rightarrow B$ and $B \rightarrow C$. On the right, the matrix represents $B \rightarrow C$ (the squares may depict different metadata, such as the support and confidence values). (Riveiro and Falkman, 2010)

(Carpenter et al. 1992) introduced a fuzzy ARTMAP algorithm originally developed by (Carpenter et al. 1991). Both works were based on the Grossberg's "Adaptive Resonance Theory" (ART) of learning in cerebral cortex. The algorithm is based on a class of neural network architecture that performed incremental supervised learning of recognition categories and multidimensional maps in response to input vectors presented in an arbitrary order. The algorithm provides fast learning that discriminates between classes of objects, events, or behaviours. (Rhodes et al. 2007), discuss learning methods and applied a modified, version of the fuzzy ARTMAP neural network classifier. This is perhaps the most heavily utilised association rule. Item-sets ascend form data linking and the discovery of the formation underpinning such data is highly important and is applied in many fields. This particular method of mining is most associated with the "Apriori" algorithm present by (Agrawal et al. 1993). An association rule is an implication of expression of the form $X \rightarrow Y$ where X and Y are disjointed item sets. An example can be stated as $X \cap Y = \emptyset$. The strength of an association rule can be measured in terms of its *support* and *confidence*. Support determines how often a rule is applicable to a given data set, while confidence determines how frequently items in Y appear in a transaction that contains X .

The formal definitions of these metrics are:

$$\text{Supports } S, (X \rightarrow Y) = \frac{\emptyset(X \cup Y)}{N}$$

$$\text{Confidence } S, (X \rightarrow Y) = \frac{\emptyset(X \cup Y)}{X}$$

A deeper association rule analysis including basic concepts and algorithms is provided by (Pang-Ning et al., 2005).

5. Mythology

Data mining techniques provide the ability to recognise previously unknown important patterns and trends. The paper presents a proof of concept for detecting anomalous vessel activity using the data mining K-means clustering method. Test simulations are based on vessel activities within the Port of Cork. The cluster data sets used for the scenario is a simulation of a 3 hour period within the Port. The marine traffic is primarily cargo vessels, Cruise liners, Tug boats and leisure craft. The activities of the vessels are given particular weights based on the latitude and longitude, speed. The port of cork is subdivided into a reference polygon based on latitude and longitude that defines the shipping lanes as indicated in figure 3. Snapshots of data were taken from the authorised shipping lane. Another data set snapshot of unauthorised vessel activity was also recorded. This activity was recorded outside the port authorised shipping zone and was not transmitting AIS.

The data used in experiments is simulated data. There are numerous fields contained within an AIS data string. However, this simulation will include only two fields, latitude and longitude. It is important to note that AIS can be manipulated and falsified, and as such cannot be classified as completely accurate; the system is only as accurate as the operator responsible for inputting data. The system is an on board ship broadcast transponder. It is compulsory for vessels over 3 tonnes to continually transmit their ID, position, course, speed, ETA and type of cargo. AIS data is transmitted to all other nearby ships and shore side authorities on VHF radio.



Figure 3: A vessels anomaly is located outside the standard shipping lane (far left circular cluster). The simulated vessel indicated no AIS data. Entered the ports restricted area and returned back to open water without port authority.

6. Framework Architecture

The system architecture consists of five main elements. (i) The Data sets – data is extracted from three data sets. These include AIS, radar, and LRIT. (ii) The framework will be based on the artificial neural network architecture (ANN) and will learning via unsupervised learning. (iii) The Human Interface – The human interface will be responsible for assessing anomalies, and updating relevant data sets.

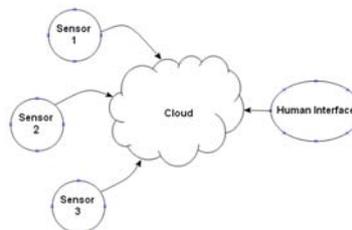


Figure 4: The architecture of the proposed framework. The processing algorithms and required hardware infrastructure is located within a private secure cloud.

6.1. Proof of Concept

A cloud instance of the frame work was implemented and tested using simulated AIS data and other vessel characteristics. The system also includes users feedback to update the machine learning surveillance model in real-time.

6.2. Data

The experimental data includes latitude and longitude co-ordinates of simulated ship movements with the port of Cork, knots, and vessel type, AIS-on/off. Noise is also added to the sensor data. Prior to testing various the clustering procedures were classified. A two-step clustering approach combining hierarchical and partitioning methods (notably the K-means procedure) were implemented. The cluster distance function was classified by the Euclidean distance; the cluster linkage type used was single. The hierarchical algorithm was used to access the number of cluster to retain from the data. This is illustrated in figure 5. Once the number of cluster was accessed the k-means algorithm was implemented. The algorithm used a setting of 5 clusters and 99 iterations.

6.3. Results

The clustering algorithm was provided with a training data set. The algorithm was set with a k means value of $k = 5$. The output provided the following data. The results demonstrate 5 distinct clusters. Figure 5 demonstrates the weights assigned to each of the clusters. The two anomalous clusters produced low percentages of 9.03% and 1.08%. The two false positives non anomalous clusters produced percentage rates of 20.94% and 34.30%. Figure 6 outline 2 anomaly clusters and 2 false positive clusters and 1 non-anomalous clusters. This is a positive results that highlights the proof of concept.

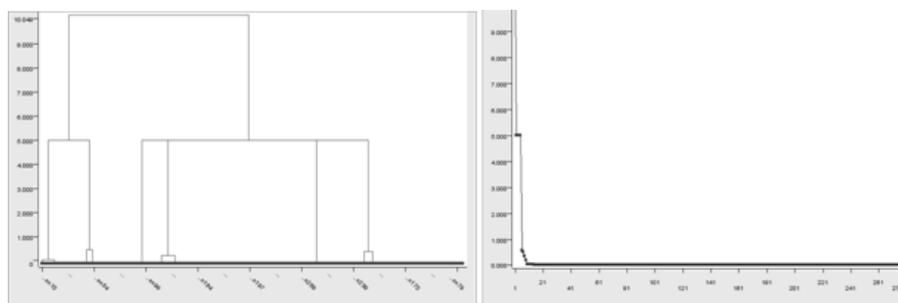


Figure 5: The Dendrogram (left) shows the hierarchical cluster structure of the data and indicates 5 clusters required for the dataset. The distance graph (right) also indicates the amount of clusters to retain from the data. The distance break (elbow) in occurs at 5.

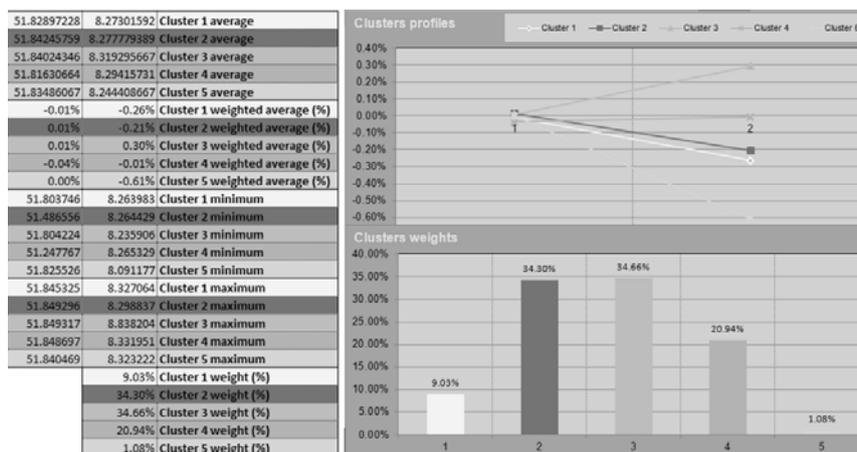


Figure 6: The results demonstrate the weights assigned to each of the clusters. The two anomalous clusters produced low percentages of 9.03% and 1.08%. The two false positives non anomalous clusters produced percentage rates of 20.94% and 34.30%.

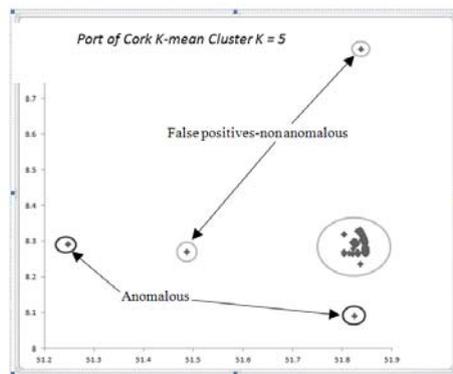


Figure 7: The scatter plot chart illustrates the clustering output. There are two anomalous clusters and two false positive clusters. The tests also produced one non-anomalous cluster.

7. Conclusions

We have introduced a general unsupervised learning framework for anomaly detection in a maritime environment. We have presented an unsupervised clustering approach that demonstrates the feasibility of cloud hosted anomaly detection for a multisensory application. Future work will focus on extending and validating this approach to unstructured data such as ship logs, historical records and social media data related to maritime activity.

8. References

- Agrawal, R., Imielinski, T. and Swani, A. (1993), "Mining Association Rules Between Sets of Items in Large Databases", *SIGMOD*, (pp. 207-216).
- Anderson, J. P. (1972), "Computer Security Technology Planning Study", *Technical Report ESD-TR-73-51*, <http://seclab.cs.ucdavis.edu/projects/history/> The Mitre Corporation, Air Force Electronic Systems Division, Hanscom AFB, Bedford, MA, 1972. Volumes I and II.
- Auslander, B., Gupta, K., & Aha, D. W. (2011), "A Comparative Evaluation of Anomaly Detection Algorithms for Maritime Video Surveillance", *Proc. SPIE 8019*, 801907; doi:10.1117/12.883535.
- Bomberger, N. A., Rhodes, B. J., Seibert, M. and Waxman, A. M. (2006), "Associative Learning of Vessel Motion Patterns for Maritime Situation Awareness", *9th International Conference on Information Fusion*, 1-8.
- Carpenter, G. A., Grossberg, S. and Reynolds, J. H. (1991), "ARTMAP: Supervised real-time learning and classification of nonstationary real-time data by a self organising neural network", *Neural Networks 4*(5), (pp. 565-588).
- Carpenter, G., Grossberg, S., Markuzon, N., Reynolds, J. H. and Rosen, D. B. (1992), "Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps", *IEEE Transactions on Neural Networks*, VOL 3, NO.5, .
- Carvalho, R. N., Haberlin, R., Costa, P. G., Laskey, K. B. and Chang, K. C. (2011), "Modeling a Probabilistic Ontology for Maritime Domain Awareness", *Proceedings of the 14th International Conference on Information Fusion (FUSION)*, Chicago, IL: PP 1 -8, 978-1-4577-0267-9 ,12177604.
- Chandola, V., Banerjee, A., & Kumar, V. (2009), "Anomaly Detection: A Survey", *ACM Computing Surveys (CSUR)*, Volume 41 Issue 3, July.
- Compieta, P., Martino, S. D., Bertolotto, M., Ferrucci, F. and Kechadi, T. (2007), "Exploratory spatio-temporal data mining and visualization", *Journal of Visual Languages & Computing*, Pages 255-279.
- Edlund, J., Gronkvist, M., Lingvall, A. and Sviestins, E. (2006), "Ruled-based situation assessment for sea surveillance", *Proceedings of SPIE Vol. 6242: Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications*. Belur V. Dasarathy, Editor, 624203.
- Endsley, M. R. (1995), "Toward a theory of situation awareness in dynamic systems", *Situation awareness, Human factors*, vol. 37, no 1, pp. 32-64.
- Fotheringham, A., & Wilson, J. P. (2008), *Geographic Information Science : An Introduction*. DOI: 10.1002/9780470690819: Blackwell Publishing Ltd.

- Guha, S., Rajeev, R., & Kyuseok, S. (2000). ROCK: A ROBUST CLUSTERING ALGORITHM FOR CATEGORICAL ATTRIBUTES. *Information Systems, Vol. 25, No 5.* , pp, 345-366.
- Hodge, V. J., & Austin, J. (2004), *A survey of outlier detection methodologies*. Artificial Intelligence Review.
- Hunter, A. (2009), "Belief Modeling for Maritime Surveillance", *12th International Conference on Information Fusion* (pp. 1926-1932). Seattle, WA: ISBN:978-0-9824-4380-4.
- Jain, A. K. (2009), "Data clustering: 50 years beyond K-means", *Elsevier, Pattern Recognition Letters*.
- Jain, A. K., & Dubes, R. C. (1988), *Algorithms for Clustering Data*. Prentice Hall.
- Jain, A. K., Duin, P. W., & Mao, J. (2000), "Statistical Pattern Recognition: A Review", *IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.22, No. 1.*,
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999), "Data clustering: a review", *ACM Computing Surveys (CSUR)*, Volume 31 Issue 3.
- Kerren, J. T., Stasko, J. D., & North, C. (2007), "Information visualization - human centered issues in visual representation, interaction and evaluation", *Information Visualization*, 6:189-196.
- Khatkhate, A. M., Ray, A. and Keller, E. (2007), "Modelling and system identification of an experimental apparatus for anomaly detection in mechanical systems", *Applied Mathematical Modelling*, pp. 734-748.
- Kimani, L. S., Catarci, T., Santucci, G. and Sartori, C. (2004), "A visual data mining environment", *Journal of Visual Languages & Computing*, 37-67.
- Kubica, J., Moore, A., Cohn, D. and Schneider, J. (2003), "Finding underlying connections: A fast graph-based method for link analysis and collaborative queries", in: T.Fawcett, N Mishra (Eds), *Machine Learning, Proceedings of the 20th International Conference, ICML 2003*, pp. 392-399.
- Laere, J. V., & Nilsson, M. (2009), "Evaluation of a workshop to capture knowledge from subject matter experts in maritime surveillance", *12th International Conference on Information Fusion*. Seattle, WA, USA, 6-9 July .
- Laxhammar, R. (2008), "Anomaly detection for sea surveillance", *11th International Conference on Information Fusion* (pp. 1-8). Cologne: Print ISBN: 978-3-8007-3092-6 .
- Munoz, T. (2011), "Maritime Domain Awareness Global Maritime Information Sharing", Retrieved January 17, 2012, from Maritime Executive: <http://www.maritime-executive.com/article/maritime-domain-awareness-global-maritime-information-sharing>
- NATO (2004 - 2012), *207 CDS 10 E bis - Maritime Security: NATO and EU Roles and Coordination*. Retrieved January 17th, 2012, from NATO Parliamentary Assembly: <http://www.nato-pa.int/Default.asp?SHORTCUT=2087>
- Pang-Ning, T., Steinbach, M., & Kumar, V. (2005), "Association Analysis: Basic Concepts and Algorithms" in *Introduction to Data Mining*, Addison-Wesley. ISBN 0321321367.
- Rhodes, B. J. (2007), "Taxonomic knowledge structure discovery from imagery-based data using the neural associative incremental learning (NAIL) algorithm", *Information Fusion* 8, 295-315.
- Rhodes, B. J., Bomberger, N. A. and Zandipour, M. (2007), "Probabilistic Associative Learning of Vessel Motion patterns at multiple spatial scales for maritime situation awareness", *Information Fusion, 2007 10th International Conference*, (pp. 1-8). Quebec, Que.
- Rhodes, B. J., Bomberger, N. A., Freyman, T. M., Kreamer, W., Kirschner, L., L'italien, A. C., et al.. (2007), "SeeCoast: Port surveillance and automated scene understanding for ports and coastal areas", in *Proceeding of SPIE Vol 6578*. Orlando, FL, USA: Defense Transformation and Net-Centric Systems.
- Riveiro, M. and Falkman, G. (2009), "Interactive Visualization of Normal Behavioral Models and Experts Rules for Maritime Anomaly Detection", *Sixth International Conference on Computer Graphics, Imaging and Visualization*.
- Riveiro, M. and Falkman, G. (2010), "Supporting the analytical reasoning process in maritime anomaly detection", *Information Visualisation (IV), 2010 14th International Conference* (pp. 26-29). London: ISSN: 1550-6037 , Print ISBN 978-1-4244-7846-0 .
- Riveiro, M., Falkman, G. and Ziemke, T. (2008), "Improving maritime anomaly detection and situation awareness through interactive visualization", *11th International conference on Information Fusion*, 1-8.
- Riveiro, M., Falkman, G., Ziemke, T. and Kronhamn, T. (2009), "Reasoning about anomalies: A study of the analytical process of detecting and identifying anomalous behavior in maritime traffic data", *SPIE Defense, Security and sensing*, 13-17.
- Riveiro, M., Johansson, F. and Ziemke, T. (2007), Supporting Maritime Situation Awareness Using Self Organizing Maps and Gaussian Mixture Models. In *10th Scandinavian Conference on Artificial Intelligence (SCAI 2008)*, pp. 84- 91.

Song , X., Mingxi, W., Jermanine, C. and Ranka, S. (2007), "Conditional Anomaly Detection", *IEEE Transactions on Knowledge and Data Engineering*, Vol 19, No 5.

Steinhaus, H. (1956), "Sur la division des corp materiels en parties", *Bull.Acad.Polon Sci. IV (C1.III)*, 801-804.

Wilkinson, L., Anand, A., & Grossman, R. (2006), "Interactive Exploration Guided pairwise Views of Point Distributions", *IEEE Transaction in Visulaization and Computer Graphics*, (pp. 1363-1372). <http://www.cs.uic.edu/~wilkinson/Publications/sorting.pdf>

Proceedings of

CERC2012

Collaborative European Research Conference

These proceedings are from CERC 2012, the Collaborative European Research Conference, which was initiated to foster research across all disciplines within European third level research and educational institutes. This issue contains high quality papers from diverse disciplines ranging from art to computer science.

ISSN : 2220 - 4164

