Modeling and Simulation of GSM/GPRS Network Derived Error-log Messages Summary using Bayesian Inference Technique

(A case study of Airtel Nig. Networks)

*1Otori, A. U.; 2Adetiba, O.E. 3Dajab, D. D. ; 4Muazu, M. B. ; 4

[1,2] Aeronautical Telecommunications Engineering School, Nigerian College of Aviation Technology, Zaria, Nigeria

[3,4] Department of Electrical & Computer Engineering, Ahmadu Bello University, Zaria, Nigeria

ABSTRACT--One of the most important requirements to be addressed by a general purpose Fault Management (FM) system is the ability to quickly identify the root cause of network errors and fix them as soon as possible. This informs the maintenance of an accurate model of the mobile network error logs for the FM task. Filtering and Correlation are two methods we used to simplify the separation of the principal alarms and redundant alarms from their side effect on network performance. An algorithm: Bayesian Inference Technique provides a platform to systematically combine the qualitative and quantitative aspects of the Bayesian model for network fault management analysis and to reduce total computational complexity by providing a database of software alarm parameters respectively. These resulted in a Global Bayesian Network which helps to represent causal chains, i.e. links between cause/effect relationships to provide the evidence of past events and predict the most likely future causes and their symptoms by computing Conditional Probabilities of each Symptom. This is critical and useful for effective Global Systems for Mobile Communications/General Packet Radio Service (GSM/GPRS) Fault Management Systems by reducing the total downtime translsting into improved quality of service (QoS)..

Keywords--: *Fault Management; Filtering and Correlation; Bayesian Inference*

I. INTRODUCTION

The complex environment in a GSM/GPRS networks has made it imperative for Service Providers to ensure more resilient network architecture and the use of object oriented software e.g. C++, java and Relational SQL Database [7] for specific tasks implemented in the OSS with minimal errors expected in the algorithm has also become important. But, network operations are usually bedeviled with all sorts of faults and network errors despite the advancement.

To help mobile service providers meet these new challenges with a better quality of service and to facilitate the smooth integration of multi-technology; the OSS has evolved into the most important element in a typical GSM network [6]. Hence, the need to ensure that network monitoring tool is provided across all network elements.

This research seeks to simplify network monitoring by deploying a powerful decision technique commonly referred to as Bayesian Inference Technique.

A. PROBLEM DEFINITION

Alarms represents symptoms and as such there are two real world concerns: The entire volume of alarm events as network software errors are logged and modelling of the causes/effects of fault types. There are no software that simulates the Cause/Effect elements of Network events error-log, Hence, an accurate model based on both Structural and empirical parameters can be used to analyse the summary of network error-log messages. Also, using derived error-log message summary of a peculiar data from a live network to validate the Bayesian Inference technique proposition:

Posterior x Evidence = Likelihood x Prior,

getting posterior probability is always difficult. Thus, resolving the Likelihood and the prior probabilities is far more easier but rarely undertaken for Fault Management task; either as a result of the difficulty in getting data or lack of database for the previous error-log messages.

B. Related work

[9] conducted a practical study at Helsink University of technology to address the need for an Automated Mobile Radio Network Performance and Fault Management System in a typical NMS (OSS).The process follows the root-cause analysis methodology by getting top level performance management data and Network error logs, and alarms trapped for Fault Management Analysis

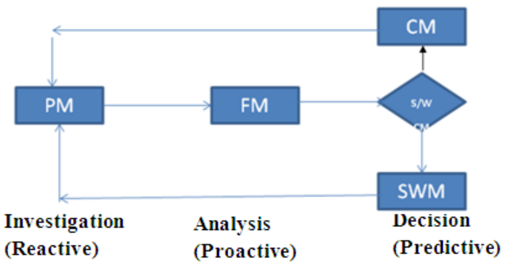

Fig. 1.0 Root-Cause Analysis Methodology

The process follows the root-cause analysis methodology by getting top level performance management data and Network error logs, and alarms trapped for Fault Management Analysis. In this work, Wallstrom stopped at Phase 1 (Investigation) to a reactive system Performance Management by analyzing some key Performance Indicators (KPI). The main thrust of this work therefore is to analyze Network error logs alarms for the Fault Management (FM) Systems which constitutes the Second phase (Analysis) to produce a proactive network system.

II. Bayesian Networks

Telecommunications networks error-logs are characterized by their structure and behavior, knowledge of these are modeled using rules that are based on known faults. These model-based systems comprises of well structured fault scenarios which can graduate from being reactive to proactive and even go on to becoming predictive in nature for fault diagnosis i.e.a form of artificial intelligence systems.

The Bayesian Model paradigm more than meets the requirements for this model-based system. Although the development of the model-based reasoning has to be completed before knowledge-based reasoning artificial intelligence systems can be developed.

Baye's Theorem stated mathematically thus:

$$P\left(\frac{H}{D}\right) = \left[P\left(\frac{D}{H}\right) * P(H)\right] / P(D)(2.3)$$

$$P(D) = \left[P\left(\frac{D}{H}\right) * P(H)\right] / P\left(\frac{H}{D}\right) \Xi \sum_i P(D_{i\,,} H_i)(2.4)$$

Where:

P(D) is the marginal probability of witnessing the data under all possible hypothesis; $\left[P\left(\frac{D}{H}\right) * P(H)\right] / P\left(\frac{H}{D}\right)$ is the joint probability that the faults occur with sets of symptoms. Hence, if we assume that the sets of alternative faults are mutually exclusive, then,$\sum_i P(D_{i\,,} H_i)$ defines the Bayesian Network model. P(D) is the normalizing constant which will be evaluated.

The Bayesian network is a pair (D, H) that allows efficient representation of a joint probability distribution over a set of random variables U= $\{X_1, \ldots\ldots\ldots\ldots X_n\}$, D is a Directed Acyclic Graph (DAG) whose vertices corresponds to the random variables X and whose edges represents direct dependencies between the variables (depicting the Causes & Symptoms).

The second component H is a set of conditional probability functions; one for each variable;

$$H = \{P(X_i / \psi_{),\ldots,} P(X_n / \psi \;)\} \quad (2.5)$$

Where; ψ is the parent set of X_i in U, the set H defines a unique joint probability distribution over U given by:

$$P(U) = \frac{\prod_{i=1}^{n} P(X_{i\,})}{\psi} \qquad (2.6)$$

In this research, the DAG will be used to encode the causal relationships between particular variables represented as nodes. Nodes are connected by causal links represented by arrows which points from parent nodes (causes) to the child nodes (effects or symptoms) and vice versa. Each node in the resulting network has associated with it a conditional probability that quantifies the effects that the parents have on the nodes [1] and [8] Taking the graph as a whole, the conditional probabilities and the structure can be used to determine the marginal probabilities or likelihood of each node holding one of its states.

Bayesian Networks are practically applied to model causal inference systems where the cause/ symptoms of faults are modeled as the nodes of a BN, while the edges represent the cause-effect relationships between these entities. The qualitative part of a BN is encoded in the structure of the diagraph and the conditional probability distribution for the nodes encodes the qualitative portion [5].

Bayesian Networks based on probability theory uses the Causal Inference and Diagnostic Inference Techniques. The causal inferences reason as a result of variables (fault types) from the given causes and the diagnostic inference reasons as a

A Simple Bayesian Network (SBN) that may be used for diagnosis in Cellular Networks, where the nodes at the edge represent the Symptoms and Causes, is shown in Fig.2.0 (a). It Consists of a parent node C,

result of symptoms only [9] and [8].

whose states are the possible Causes, and children nodes S_1 ……… $S_{m,}$ Fig.4.1 (b) depicts the Simple Baye's Model which represents the symptoms and may have any discrete number of states [3]

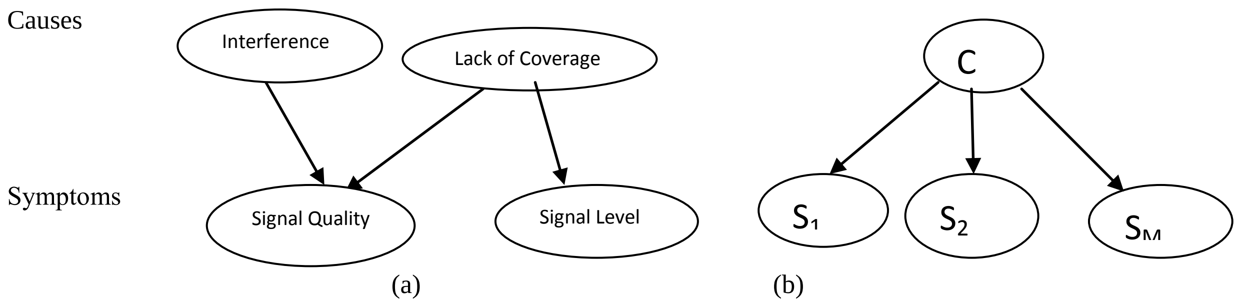

Fig. 2.0 Example of Bayesian Network for Diagnosis in Cellular Networks and Simple Baye's Model

III. Software Alarm Filtering and Correlation Techniques

Quantitative and qualitative models can be used to predict which components are likely to contain the highest concentration of faults based on adequate software metrics and log of faults found by testers and clients of software systems. To develop such systems, a complete understanding of Network Management principles is required. This is valid for mobile radio networks where an efficient fault management system should reduce the outage time of radio and other communicating resources is achieved by means of an automated analysis of the alarms generated and by an automated diagnostic process. It is based on results of the alarm filtering correlation that a fault diagnosis is made; this can vary from simple message filtering and redundant alarm suppression to a more sophisticated alarm compression and generalization techniques [3][4].

IV. METHODOLOGY AND DATA COLLECTION

A. Methodology

- Network Elements (NE) Software alarms and error logs i.e. overall network error- log message from NEs like, BSC and MSC of Airtel Network was collected for a period of Three (3) Months.
- A method called 'Tupling' was used to filter the raw and voluminous error log messages at both the BSC and MSC. The Filtering and Correlation will be realized by writing Advance sorting and grouping program using the Excel Formulae.
- Bayesian modelling was used to model error-log message summaries.
- Formulate our knowledge of the situation probabilistically by defining the model that expresses quantitative aspects of the known parameters using the following expressions from Bayesian Theorem to specify the marginal probabilities and the conditional probabilities of known fault types given the cause(s) respectively,

$P(U) = \prod_{i=1}^{n} p(\frac{Xi}{\psi})$,

P (S_i=$S_{i,j}$ /C=C_k)

where, ψ is the parent set of X_i in U,

the set H defines a unique joint probability distribution over U, X_i

($S_{i,j}$) is the random variable i.e. the symptoms, and C_k is the prior probability of the cause.

- A standard Conditional Probability Table (CPT) of fault type's conditional probability distribution was computed by dividing the appropriate marginal.

B. DATA COLLECTION AND PRESENTATION

The data used in this work was obtained from AirTel Networks Network Management Center, Lagos. Error log files of network elements like: MSC and BSS (BSC) consisting of error log-Messages was collected for a period of three (3) Months (January, 2010 to March, 2010);

V. RESULTS AND ANAYSIS

Results are presented graphically so that Fault Trend Analysis and Fault Trees of error –log messages showing variations/behaviour of one measured parameter against another is obtained and inferences used to draw logical conclusions.

Also, MATLAB Simulation of the relational database for error –log messages which help to provide a database of previous faults for the purpose of analysis is realized

A. BAYESIAN MODELING OF BSS AND MSC DERIVED ERROR-LOG MESSAGE SUMMARY

A Bayesian network model has two components and is a pair: (D, H). ''D''is a Directed Acyclic of a BN, while the edges represent the cause-effect relationships between these elements. The qualitative part of a BN is encoded in the structure of the diagraph and the conditional probability distribution for the nodes encodes the quantitative structure. The Bayesian Model consists of two phases, Phase one is the qualitative aspect while phase two will encode the quantitative aspect.

B. QUALITATIVE MODELING OF NETWORK ELEMENTS ERROR- LOG MESSAGE SUMMARY USING SIMPLE BAYESIAN MODEL (FIRST PHASE)

A careful look at the Conditional Probability Table of BSS Alarm summary shows that for each

To aid in the quest for an automated diagnosis of network alarm's fault management system; the Bayesian model can also be used to represent the faults occurring at the MSC in a Diagraph. The fault types that were collected and filtered followed by correlation exercise are assigned new names i. e. they are masked. The alarm summary in Appendix 2B is in agreement with the fact that many fault types (symptoms) could be triggered by a number of causes or just a single cause.

A critical look at the table of Appendix 2B shows that there are three major causes of faults: Equipment (with all the hidden causes); Transmission network problems and Software error. Taking the causes and the symptoms as the elements of our Simple Bayesian Model gives the modeled fault tree for classified MSC Alarms shown in Fig. 4.0 This depicts the qualitative structural Bayesian model for MSC Fault summary. It is obvious from the representative model that the number of triggered alarms for a single cause is normally very high. In addition, the same alarms may be triggered by different cause. The Simple Bayesian Model helps to represent the well structured fault classes using the two most

Graph (DAG) whose edges represent direct dependencies between the variables i.e. the Causes and Symptoms as nodes and the elements of the qualitative aspect of the Model, while ''H''is a set of conditional probability functions; one for each variable and this represents the quantitative aspect of the model. One specific feature of a mobile communications Fault Management Systems is that: a fault cause can result in a large set of similar symptoms, as there is no single fault assumption built into the process, hence, each possible explanation can be a conjunction of single cause.

Bayesian Networks are practically applied to model causal inferences systems where the modules of a system are modeled as the nodes

fault caused, there are a number of symptoms that are logged at the NMC. The error log summary agrees with the assertion that a single fault cause results in a number of symptoms. This constitutes the first phase of our modeling (Fig.3.0).There are only two causes of fault symptoms at the BSS (Appendix), i. e. Equipment and Transmission Network and each of this fault causes resulted in a number of symptoms (Fig. 3.0). This simple model helps a lot in drawing up commonalities between the faults which further takes us steps closer to the much desired well structured fault model for automatic diagnosis. Fig. 4.0 shows the Simple Bayesisn Model of the derived Error-log message at the MSC

important elements in fault analysis which are the symptoms and their root causes.

C. QUANTITATIVE MODELING OF NETWORK ELEMENTS ERROR- LOG MESSAGE SUMMARY (SECOND PHASE)

The Bayesian network model has associated with it a Conditional Probability Table (CPT) that quantifies the effects that the parents have on the nodes depicting the quantitative aspect of the model. The graph as a whole, the conditional probabilities and the structure can be used to determine the marginal probabilities or likelihood of each node holding one of its states.

The second component (H) is a set of conditional probability functions, one for each variable (as Symptoms), that is:

$$\boldsymbol{H} = \{\boldsymbol{P}(\boldsymbol{X_i}/\boldsymbol{\psi})\ldots\ldots\ldots\{\boldsymbol{P}(\boldsymbol{X_n}/\boldsymbol{\psi})\}_{(4.1)}$$

This is equivalent to a unique joint probability distribution over U represented as:

$$P(U) = \prod_{i=1}^{n} P(X_i / \psi_i)_{(4.2)}$$

This constitutes the second phase of modeling which consists in defining the parameters for the CPT.

The required probabilities to completely define the model are:

1. The prior probabilities, $P(C=C_k)$ for each fault causes C.
2. Conditional Probabilities of the Symptoms given the Causes, $P(S_i=S_{i,j} / C=C_k)$ for each state X.

There are only two causes for all classes of faults in the BSS (Table I): Equipment and transmission, the probability $P(C=C_k)= 0.5$ was specified as uniform priors. A slight deviation from this assumption is necessary when a particular fault types occur many times before others are logged. Hence, the prior probability of caues for RSL (which occurs 1100 times out of 2096 error-logs) is slightly adjusted to 0.6. Also, there are three (3) causes for all fault types occurring at the MSC (Table I): Equipment, Transmission, & Software errors, with Twenty Six (26) symptoms in all and Equipment caused about 16 of such fault types as the dominant cause i.e. P(C1=16/26) = 0.6. The remaining Ten 10 symptoms were either caused by Transmission & S/W errors, i.e. P(C2= 10/26) = 0.4. This shows $\sum Pc = 1$, and since Bayesian modeling recommends the use of uniform priors, hence, the uniform prior of 0.5 P (C_k = 0.5) is used here as well.

The CPTs derived from the BSS and MSC alarm summaries are presented in Appendix 3 (A and B). The parent set of X_i isψ = 2096, i.e. the total number of times error messages are logged and C_k = 0.5 for the BSS while the Parent set of X_i is ψ =6143 for the MSC..

The systematic combination of the Structural model and numerical data of the overall fault (alarm) summaries for both BSS and MSC can now be realized in a Global Bayesian Network (GBN) as shown in Fig 3.0 and 4.0.

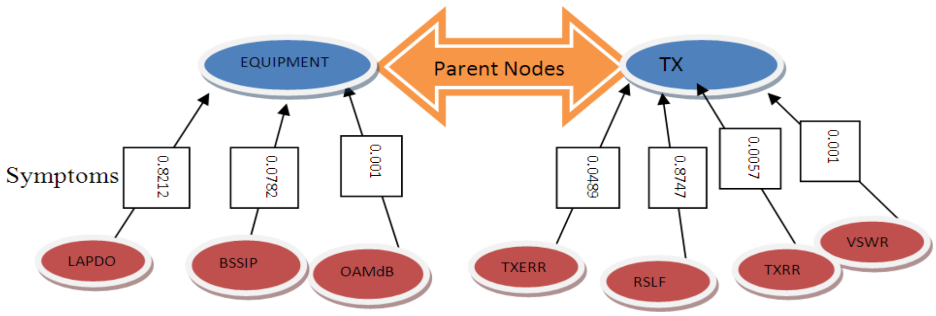

Fig. 3.0 Global Bayesian network model for BSS Error-Log Summary

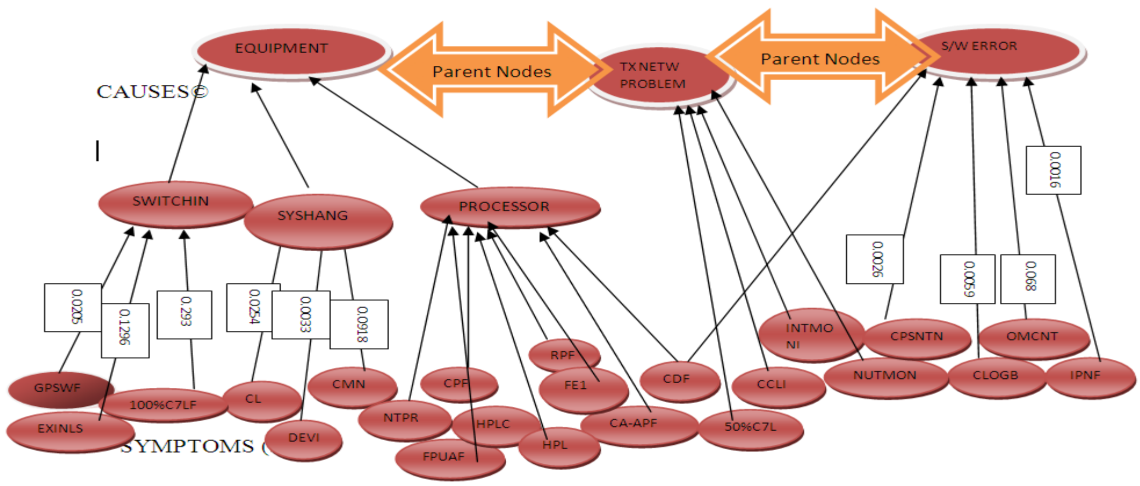

Fig. 4.0 Global Bayesian network model for MSC Error-Log Summary

Table 1
(A) Conditional Probability Table of BSS Alarm summary

Fault Type (Symptoms)	***Root Cause C***	***Frequency (No. of occurences (X_i)***	***Marginal probability*** $\frac{P(X_i}{\psi_j)} = Si,j$	***Conditional ProbabilityOf S given C*** $P(S_i=S_{i,j} /C=C_k)$
LAPD_OML failure	*Equipment Radio*	*861*	*0.4106*	*0.8212*
RSL link Failed	*Transmission µW*	*1100*	*0.5248*	*0.8747*
Transmission Network Error	*Transmission µW*	*46*	*0.0295*	*0.0489*
BSS Internal Problems	*Equipment Radio*	*82*	*0.0391*	*0.0782*
VSWR fault	*Transmission*	*1*	*0.0005*	*0.001*
Transmission rerouting	*transmission Mux*	*6*	*0.0029*	*0.0057*
OAM Database not Found	*Equipment server error*	*1*	*0.0005*	*0.001*

D. INTERPRETATION OF RESULTS

The Global Bayesian Network model graphs or fault trees is considered to represent the joint probability distribution for all the variables as depicted (Fig. 3.0 and 4.0), each node has associated with it a conditional probability that quantifies the effects their parents (C) have on the nodes S. Taking the tree as a whole, the conditional probabilities and the structure can be used to determine the marginal probability or likelihood of each parent holding one of its states(S). This simply means that the Directed Acyclic graph (DAG) so derived is singly connected, each link is a bridge from symptom to root cause, where the change of one link leads to a different conditional probability due to new update and new faults will create an entirely new link.

Two components of the automatic diagnosis system may be distinguished: the model and the inference model. The model represents the knowledge on how the identification of fault causes is carried out given the symptoms; while the inference method is the algorithm that identifies the cause of the problems based on the value of the symptoms and the severity based on their value between 0 and 1as can be seen from the CPT (Table I).

Talking about the Inference systems, there is causal inference (as in the Cause –effect phenomenon) and the diagnostic inference (using the BN Probabilistic modeling).The peculiar data in this case; thus serves as an Inference Algorithm which can be used to infer that:

$P(X_j = x/E)$, represents the probability of a symptom X_j, being in certain state x, given the available evidence, E (cause), as a normalizing constant (thus informs the use of uniform priors P(C)). Off course, the target probability of the symptoms given the Cause is stated thus:

$P(S_i=S_{i,j} /C=C_k)$, which satisfies the unique joint probability function:

$$P(U) = \prod_{i=1}^{n} P(X_i \quad /\psi_{i\,)}\,; \qquad \min 0, \quad \max 1$$

Now, if the product rule (from Baye's theorem), is applied to the sets of conditional probabilities, approximate values taken as 0 means the fault is almost non-existent and 1 is a case of constantly reoccurring multiple alarm for a particular error log message, e.g. RSL link failed with a target conditional probability of approximately 1.

As the value of the conditional probabilities varies from 0 to 1; the exact value is an indication of the severity of the particular fault type (Symptom). The Inference algorithm varies in much the same respect, they are either exact; approximate or heuristic values on singly or multiply connected graphs and are used for different inference task. The causes and symptoms for fault diagnosis can be modeled as random variables with two states: absent/ present; Off/On and as discrete random variables-[0,1]. It should be pointed out that since a random variable can only be in one of its finite states at a time, it is assumed that there is only one fault occurring at a time which is not always true in real life, hence the superiority of the Bayesian Inference diagnostic algorithm over Causal Inference Technique .

VI. CONCLUSIONS

This research has validated by the simple bayesian model of error log summary that a single fault can result in multiple symptoms and a single alarm may be triggered by different causes. Furthermore, the explicit combination of the quantitative and qualitative aspects(cause and symptoms) of a network model in a Global Bayesian network gives the representative Directed Acyclic Graph showing the causal chains to supply evidence of past event to see what the most likely cause of future errors might be.

REFERENCES

[1] Baras, J.S. and Mykoniatis, H. G. (1998, April 3). *Integrated Distributed Fault Management for Communications Networks.Technical Report, CSHCN TR 98 – 10, p. 1 – 4.*

[2] Barco, R. et al. (2007). " *Automatic Diagnosis of Mobile Communications Networks Under Imprecise Parameters*. Sciencedirect Elesevive Journal,

[3] Boix, C. and Svedberg, M. (2008). *Model based Alarm Coordination system for primary and secondary alarm notifications*. Verizon business network services Inc., Patentdocs, Arlighton, VA, USA. Retrieved January 20, 2011 from Patendocs database.

[4] Carls, G., Nedil, W., Tuchis, K. (1997). *Using neural networks for alarm correlation in cellular phone networks. In Proceeding of the IEEE International Workshop on applications of Neural Networks to Telecommunications, Melbourne, Australia*

[5] Cooper, G.(1990). *The Computational complexity of probabilistic inference using belief Networks.In proceeding of IEEE on Artificial Intelligence, Milano. 42(10), 393 – 405.*

[6] Dereviere, Van M. and Verbuas, J. (2003).*Common GSM software components. Alactel Telecommunications Review*.21(1), 124-128..

[7] Hughes, S. (2006). Enterprise architecture for next generation telecommunications service provider. *Huges Technical Journal*.Retrieved October 12, 2010 from http://www.hsc.com/

[8] Lauritzen, L. and Spiegelhalter, D. (1998). *Local computations with probabilities on graphical structures and their applicationa to expert systems. J. R. Statistic Soc. B. Malaga, Spain. 50 (2), 157 – 224.*

[9] Wallstrom, M. (2007). *Automation of Mobile Radio Network Performance and Fault Management*. MSC Thesis, Helsinki University of Technology. Retrieved June 7, 2010 from the Department of Electrical Engineering online database.www.