Fault analysis and prediction in telecommunication access network

Conference Paper · February 2000

DOI: 10.1109/MELCON.2000.880386 · Source: IEEE Xplore

Damir Medved
Ericsson Nikola Tesla d.d.

Dean Saric
Klinička bolnica "Sveti Duh"

Fault Analysis and Prediction in Telecommunication Access Network

Damir Medved
Koraljka Brlas
damir.medved@ht.hr
koraljka.brlas@ht.hr
HT- Croatian Telecomm
Telecommunication Center Rijeka
HR-51000 Rijeka, CROATIA

Davor Šarić

davor.saric@ht.hr

HT - Croatian Telecom
Non-voice Services Division
HR-10000 Zagreb, CROATIA

Abstract

Technological innovations and growing consumer demand have led to a variety of design and exploitation problems in telecommunication networks. In particular, local access networks have received a lot of attention, since they account for approximately 60% of total investments in communication facilities. Hence even marginal improvements in the access network design and maintenance can lead to substantial savings.

Faults in telecommunication network have direct impact on its availability and maintenance costs, so their quick elimination, prevention and removal of causes that generated them, is of special interest. Possibility to analyse spatial distribution of faults and to predict places where future errors may arise can significantly help telecom operators who are in charge to detect and repair such problems. This data is also crucial for future planning and design of telecommunication networks as it can help to avoid problematic areas or use additional measures for protection of telecommunication cables and equipment.

*Existing statistical data (fault data and foreign factors data like precipitation, high-voltage installations etc.) is analysed and then a model of faults in telecommunication network is created using Bayesian methodology. The whole procedure of acquisition and interpretation of data is described, and an application called **FATAN** (Fault Analysis in Telecom Access Network) that aggregates all procedures is created.*

Achieved result gives us possibility to successfully determine locations where is likely that an error will arise.

1. Introduction

In past primary goal for most telecom operators in developing countries like Croatia was to reach all potential customers with telecom infrastructure, and do it fast. Shortage of adequate funding for replacing all older networks, leaded to weary heterogeneous situation (cables old up to 40 years in urban areas, cable insulations from paper to PVC, older networks with average lengths of 2,5 – 4 km, newer networks up to 1 km, etc.). Maintaining such a network is quite complex problem; therefore an adequate documentation and fault analysis system is absolutely necessary.

Today hottest question among telecom operators is QoS (Quality of Service) and how to satisfy customers and maximize profits. If you already have building blocks (applications, databases) such a described system, the logical next question is can you use this existing data to predict locations of future faults (with acceptable margins of probability) and prevent or minimize network downtime.

The main purpose of this paper is not to discuss the detailed modelling issues of faults in local access network, rather our emphasis is to show design and implementation of the procedures in the fault tracking and analysis system that may become a standard building block of the company decision support system.

2. Problem Description

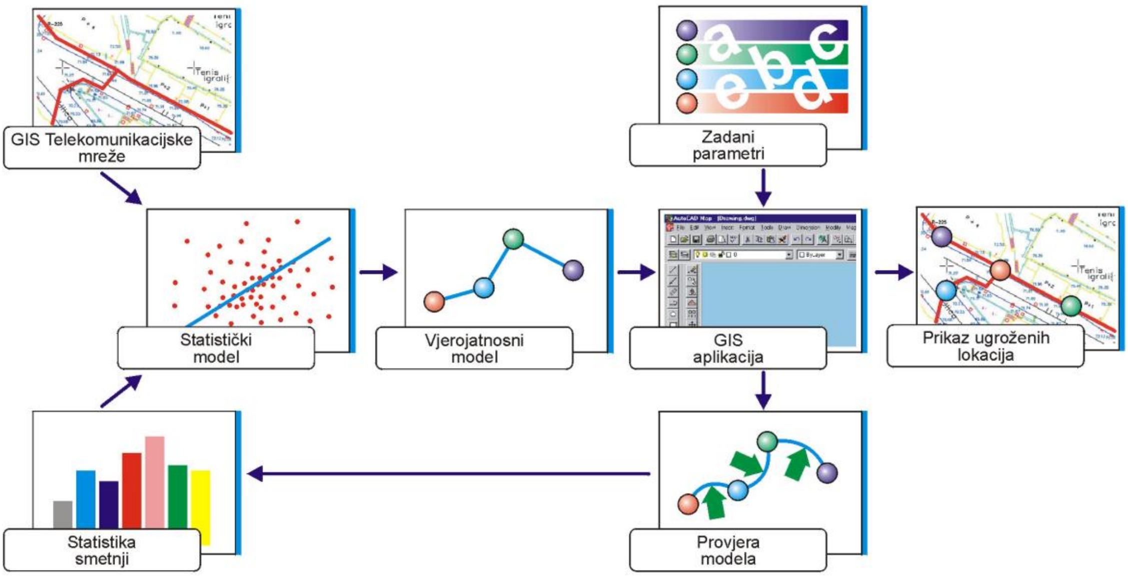
When problem was initially defined the first assumption was that the simplified model of telecom network would be used. There are many reasons for that; real world telecom network consists of huge number of network elements, which are built with many smaller parts and so forth. Obviously, description of all that elements (and all possible locations and sources of faults) is practically impossible. The relatively small Rijeka area has more than 140 000 subscribers connected with more than 800 cables to 200 switches this leads to few million possible locations of faults. Next problem is that actual system for fault tracking does not acquire in many cases lot of details (for example exact location of fault ((coordinates)), description of fault is somewhat arbitrary, for some networks technical documentation is not complete or actual, etc.).

The main sources of data are along with mentioned fault-tracking system are Telecom GIS system and meteorological data

GIS database currently holds site and schematic cable diagrams, duct and manhole drawings as well as OSP (DP, CCP) and switching equipment location and capacity. Georeferenced building and equipment locations with fault data attached are available for analysis. Up to now primary purpose of the GIS in TKC Rijeka was to support network operations and maintenance. Practically all departments use this database, primarily for real world visualization of telecom infrastructure, localization of demand for new lines, equipment and cable failure troubleshooting.

When the GIS database was created, the primary goal was to establish connection between alphanumerical data in SQL DBMS and georeferenced information contained in numerous AutoCAD schematic and site drawings. In the future, special attention will be paid to adopt data acquisition and preparation procedures that will emphasize the importance of collecting new details in addition to O&M related data, and this is the direct consequence of the development and implementation of described system.

Along with telecom data GIS database contains data from local power company (high voltage installations, transformers), water installations, flood data etc. External meteorological database contains detailed data (daily acquisition) from many locations with temperature and precipitation.



Picture 1. General model of fault analysis in telecommunication access network

Generally all process of analysis can be dissolved in few steps (Picture 1.):

1. From GIS System extract locations that are by definition more threatened (near high voltage equipment, high areas with greater possibility for atmosphere discharges, flood areas etc.)
2. Perform an statistical analysis of related fault data from fault tracking system
3. Perform statistical analysis of meteorological data and other external factors
4. Check if there is statistically significant correlation between points 1,2,3
5. Define graphical probability model (Bayesian network)
6. And finally, extract and visualize menaced objects using GIS application

3. Tools

As mentioned before first step in process of data analysis is data preparation and transformation in form suitable for further work. In our case, GIS database is build using standard AutoCad Map 3.0 and ESRI ArcView 3.1

tools. Statistical and belief network analysis is done using Netica Application 1.12 and Netica Programmer's Library, also known as the "Netica API" (Application Program Interface).from Norsys. Netica is a versatile, fast, user-friendly program that can be used to find patterns in data, create diagrams encoding knowledge or representing decision problems, use these to answer queries and find optimal decisions, and create probabilistic expert systems.

4. Bayesian networks

A Bayesian network is graphical model for probabilistic relationships among a set of variables. Few years ago researches have developed methods for learning Bayesian networks from data. These techniques are new and still evolving, but they have are quite effective for some data-analysis problems. There are four main reasons why use Bayesian methodology in fault prediction system [1]:

1. Bayesian networks can readily handle incomplete data sets. For example, in classification or regression problem where two of the explanatory or input variables are strongly anti-correlated. This correlation is not a problem for standard supervised learning techniques, provided all imputes are measured in every case. When one of the inputs is not observed, most models will produce inaccurate prediction, because they do not encode correlation between the input variables. Bayesian networks offer a natural way to encode such a dependencies.

2. Bayesian networks allow learning about casual relationships. This is important for at least two reasons. The process is useful when we are trying to gain understanding about a problem domain, for example, during exploratory data analysis. In addition, knowledge of casual relationships allows us to make predictions in the presence of interventions.

3. Bayesian networks in conjunction with Bayesian statistical techniques facilitate the combination of domain knowledge and data. Bayesian networks have a causal semantics that makes the encoding of casual prior knowledge particularly straightforward, and also encode the strength of casual relationships with probabilities. as consequence, prior knowledge and data can be efficiently combined.

4. Bayesian methods in conjunction with Bayesian networks and other types of models offers an efficient and principled approach for avoiding the over fitting of data. This means that models can be "smoothed" in way that all available data can be use d for training.

4.1. Belief network learning

Belief network learning [2] is the automatic process of determining a suitable belief network, given data in the form of cases (Picture 2). Each case represents an example, event, object or situation in the world (presumably that exists or has occurred), and the case supplies values for a set of variables, which describe the event, object, etc. Each variable becomes a node in the learned network (unless you want to ignore some of them), and the various values it takes on become that node's states. Some cases may not have values for some variables that other cases do, which are known as missing data (shown as * in table).

TK mreža 1/97 - 5/99 : Table

promet	zastita	prom_tempera	aparat	instalacija	radovi	tip_kabela	razdjelnik	izvod	lok
Ne	Nikakva	Velike	Da	Da	Ne	Olovni	Ne	Da	*
Ne	*	Male	Ne	Ne	Da	Olovni	Ne	Da	*
*	*	*	Ne	Da	Ne	*	Da	Ne	*
Ne	*	Velike	Ne	Ne	Da	Plastièni	Ne	Da	*
*	*	*	Ne	Da	Ne	*	Ne	Ne	Van
*	Nikakva	*	Da	Ne	Ne	*	Da	Da	*
Ne	Nikakva	Velike	Ne	Da	Ne	Olovni	Ne	Da	*
*	*	*	Ne	Da	Da	*	Ne	Da	*
*	*	*	Ne	Da	Ne	*	Ne	Ne	*
*	*	*	Da	Da	Ne	*	Ne	Ne	*
*	*	*	Ne	Da	Da	*	Ne	Ne	Unu
*	*	*	Da	Da	Da	*	Ne	Ne	*
*	*	*	Da	Da	Ne	*	Ne	Ne	*
*	*	*	Da	Ne	Da	*	Ne	Da	*
*	*	*	Da	Da	Da	*	Ne	Da	*
Ne	*	Velike	Da	Ne	Da	Olovni	Da	Da	*
*	*	*	Ne	Da	Da	*	Da	Ne	Van
*	*	*	Da	Ne	Ne	*	Ne	Da	*
Ne	*	Velike	Ne	Ne	Da	Olovni	Ne	Da	*
*	*	*	Ne	Da	Ne	*	Ne	Da	*
Ne	Nikakva	Velike	Da	Da	Ne	Olovni	Ne	Ne	Unu
Ne	Slaba	Velike	Ne	Ne	Ne	Olovni	Ne	Da	*
Ne	*	Velike	Ne	Ne	Da	Armirani	Ne	Da	*
*	*	*	Da	Ne	Da	*	Ne	Da	*
Ne	*	Velike	Da	Da	Ne	Olovni	Ne	Ne	Van

Picture 2. Case file for network learning

The learned network can be used to analyze a new case drawn from a population similar to the cases used for learning. Typically the values for some variables of the new case will be known. These are entered as findings, and then probabilistic inference is done to determine beliefs for the values of the rest of the variables for that case. If the links of the network correspond to a causal structure, and the query nodes are ancestors of the findings nodes, then you could say that the network has learned to do diagnosis. If the query nodes are descendants, then the network has learned to do prediction, and if the query node corresponds to a "class" variable, then the network has learned to do classification. Of course the same network could do all three, even at the same time.

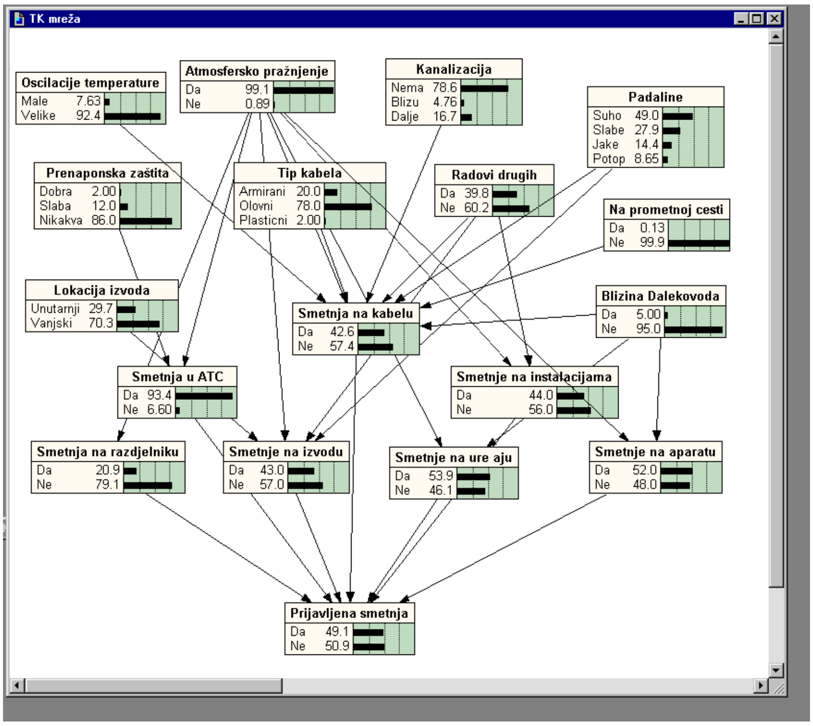
The belief network-learning task has traditionally been divided into two parts: structure learning and parameter learning. Structure learning determines the dependence and independence of variables and suggests a direction of causation, in other words, the placement of the links in the network. Parameter learning determines the conditional probability relationship at each node, given the link structures and the data. Currently Netica only does parameter learning (i.e., you link up the nodes before learning begins).

Netica uses a true Bayesian learning algorithm. While learning, Netica assumes independence between each of the conditional probabilities of a node's relation to its parents. This works very well when there is lots of data (or nodes have few parents), but can result in under-confidence and poor generalization otherwise. Netica may not have seen some parent configuration, so the learned probability distribution is uniform for it, but Netica might do better if it assumed that parent configuration behaved similarly to other similar parent configurations (i.e., assumed a little more independence between variables, and a little less independence between conditional probabilities)

5. Computational Results

In analysis of telecom access network fault data for Rijeka region more than 65000 cases where considered (from years 1997 to 2000). This is relatively short period of time but it was impossible to collect all necessary data for years before 1997. As new data is constantly acquired the model with self enhance in future, and this is probably one of the most valuable achievements of this approach.

Using Netica [4], a belief network for faults in telecommunication networks is constructed which captures believed relations (which may be uncertain, stochastic, or imprecise) between a set of variables which describes simplified model of telecommunication network (Picture 3.).When the belief network is constructed, one node is used for each scalar variable, which may be discrete, continuous, or propositititional (true/false). .

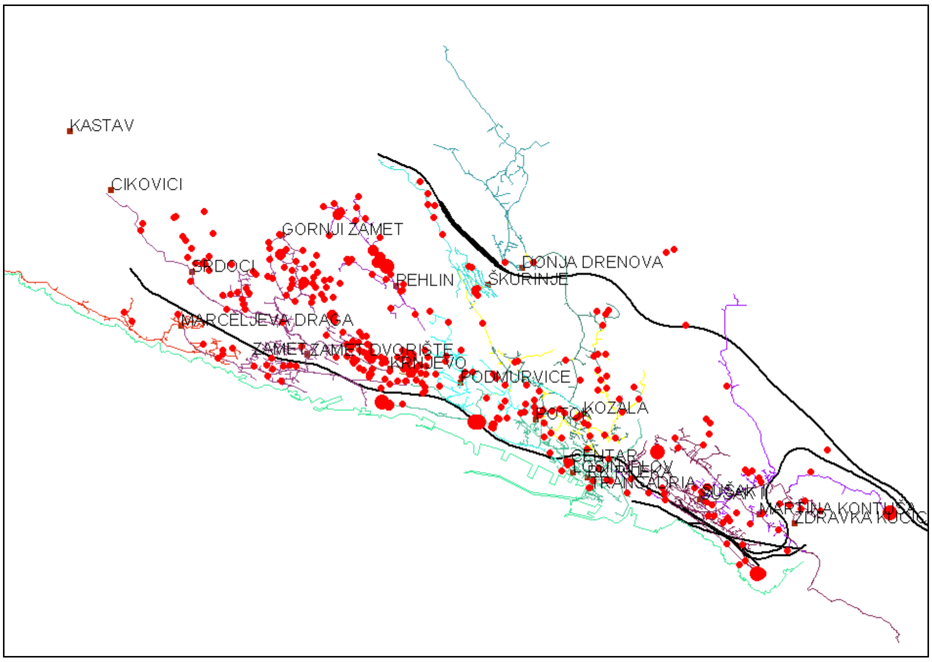


Picture 3. Belief network for telecom fault prediction

The nodes are then connected up with directed links. If there is a link from node A to node B, then node A is sometimes called the parent, and node B the child (of course, B could be the parent of another node). Usually a link from node A to node B indicates that A causes B, that A partially causes or predisposes B, that B is an imperfect observation of A, that A and B are functionally related, or that A and B are statistically correlated.

Finally, probabilistic relations are provided for each node, which express the probabilities of that node taking on each of its values, conditioned on the values of its parent nodes. Some nodes may have a deterministic relation, which means that the value of the node is given as a direct function of the parent node values.

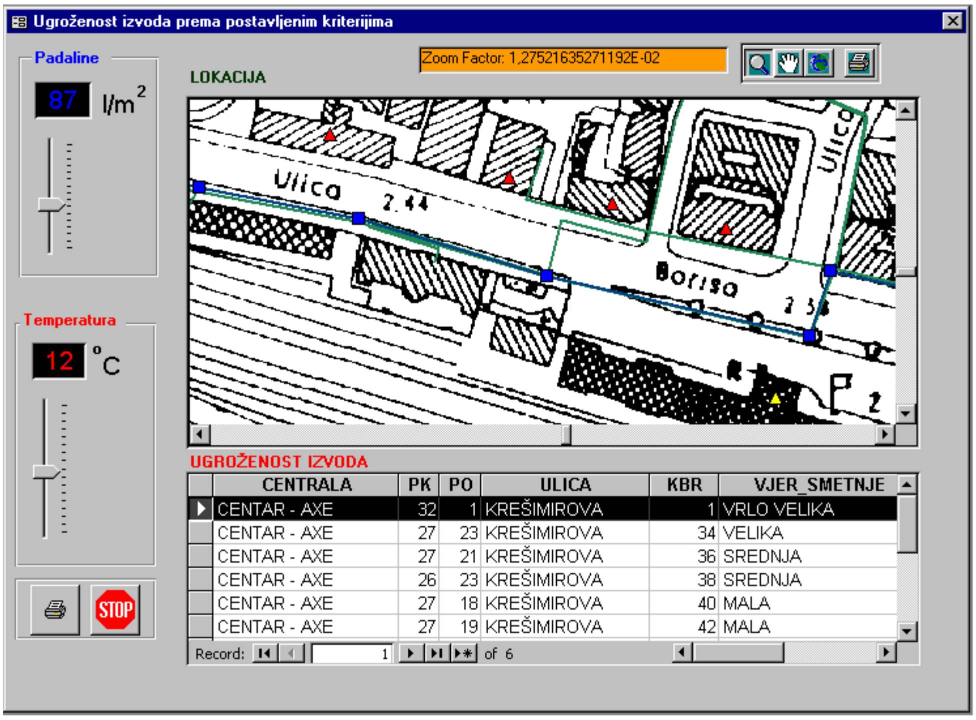
After the belief network is constructed, it may be applied to a particular case. For each variable you know the value of, you enter that value into its node as a finding (also known as "evidence"). Then Netica does probabilistic inference to find beliefs for all the other variables. Suppose one of the nodes corresponds to the variable "Temperature", and it can take on the values cold, medium and hot. Then an example belief for temperature could be: [cold - 0.1, medium - 0.6, hot - 0.3], indicating the subjective probabilities that the temperature is cold, medium or hot. Changing values (temperature, precipitation etc.) of variable or variables we can obtain probabilities of fault in given telecommunication network and after visualization with GIS system we can easily locate problematic areas (Picture 4.)



Picture 4. Example of fault analysis for variable electric discharging (greater spots define most problematic locations) – analysis show that most problematic areas are near by large metallic structures (high voltage transformers, oil raffinery, etc.)

6. Overall System Design

The FATAN system is implemented in Microsoft Visual Basic 6.0 and ESRI MapObjects 2.0. It has user-friendly interface, and it is possible to choose various types of analysis, choose area and telecommunication equipment etc. As sample (Picture 5), a module that allows changing variables for precipitation and temperature and immediately shows probabilities for faults in selected area, is given.



Picture 5. Probability of fault when changing variables for temperature and precipitation

7. Further Work

Greatest problem in fault prediction in telecommunication access networks is missing or incomplete data, so greater effort for modifying and enhancing existing applications for fault tracking and enforcing discipline in data acquisition will be primary goals in years to came.

Tighter integration of FATNET with the GIS is a natural next step that will provide more efficient analysis environment. In addition to that, new forecasting modules will enhance overall capabilities of system.

8. Summary and Conclusions

The three pillars of the project;

1. Implementation of an efficient and proven Bayesian methodology to problem of fault analysis in telecommunication networks,
2. Usage of all available data stored in many different telecom and external databases, efficient conversion and aggregation of various types of data and enabling fast problem modeling and visualization,
3. Up-to-date fault analysis and simulations from the modern GIS application,

has enabled a small team of professionals to achieve valuable results in a short available time, and show that is possible and economically rightly build such a system, all in desire for further enhance QoS for our customers.

References

[1]. David Heckerman et al.: *"Learning Bayesian networks: The combination of knowledge and statistical data"*, Machine Learning 20, pages 197 – 243, 1995

[2]. David Heckerman: *"A Tutorial on Learning With Bayesian Networks",* Technical Report, Microsoft Research – Advanced Technology Division, pages 2-5, November 1996

[3]. David Heckerman: *"A Bayesian approach for learning casual networks",* Proceedings of Eleventh Conference on Uncertainty in Artificial Intelligence, pages 285 –295, Montreal, 1995

[4]. Netica 1.12 User's Guide, Norsys Inc., 1999