

# Application of probabilistic modeling and machine learning to the diagnosis of FTTH GPON networks

Stephane Gosselin, Jean-Luc Courant, Serge Romaric Tembo Mouafo,  
Sandrine Vaton

## ► To cite this version:

Stephane Gosselin, Jean-Luc Courant, Serge Romaric Tembo Mouafo, Sandrine Vaton. Application of probabilistic modeling and machine learning to the diagnosis of FTTH GPON networks. ONDM 2017: 21st Conference on Optical Network Design and Modeling, May 2017, Budapest, Hungary. pp.1 - 3. hal-01573963

**HAL Id: hal-01573963**

**<https://hal.archives-ouvertes.fr/hal-01573963>**

Submitted on 11 Aug 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Application of Probabilistic Modeling and Machine Learning to the Diagnosis of FTTH GPON Networks

Stéphane Gosselin, Jean-Luc Courant  
Orange Labs  
Lannion, France  
stephane.gosselin, jeanluc.courant@orange.com

Serge Romaric Tembo  
ActiveEon  
Sophia Antipolis, France  
serge.tembo@activeeon.com

Sandrine Vatou  
IMT Atlantique  
Brest, France  
sandrine.vatou@imt-atlantique.fr

**Abstract**—This paper presents insights on the promises of probabilistic modeling and machine learning for fault diagnosis in optical access networks. A Bayesian inference engine, called Probabilistic tool for GPON-FTTH Access Network self-Diagnosis (PANDA), is applied to fault diagnosis of Gigabit capable Passive Optical Networks (GPON). PANDA approach has been assessed on real diagnosis data, showing very satisfactory alignment with an operational rule-based expert system. Furthermore, it provides diagnosis conclusions for all tested cases, even if some monitoring data is missing or incomplete. Finally, an expectation maximization algorithm allows to finely tune the probabilistic model.

**Keywords**—*Network management; optical access network; fault management; model-based approach; Bayesian inference; machine learning; expectation maximization*

## I. INTRODUCTION

Significant efforts are made by telecommunication operators to provide broadband services with high availability. This includes improvements of reliability and robustness of network infrastructures, through automation of complex network management operations such as fault diagnosis [1].

Fault diagnosis aims at detecting and locating any type of failures that degrade the quality of service or lead to service unavailability. A fault occurrence typically triggers other faults and alarms, which might result in complex and distributed patterns of failures and alarms. In particular, a given fault may lead to several alarms, and a given alarm may result from various faults. Fault diagnosis consists in finding the root causes which have triggered a given set of alarms in specific network conditions. The latter are derived from monitoring data which is gathered by the network management system.

Fault diagnosis requires expert knowledge which is typically involved in designing rule-based expert systems [2]. As a matter of fact, the current diagnosis tool used by Orange for its FTTH networks relies on an expert system based on a

decision rule engine. This rule-based system reaches high diagnosis performance, but maintenance and creation of new rules is difficult and requires the analysis of large amounts of data. Due to the dynamic, distributed and non-deterministic nature of networks, fault propagation is a complex phenomenon and available rules do not cover all possible fault configurations. This leads to situations in which the expert system is unable to diagnose the root causes of network faults.

Other diagnosis methods reproduce fault and alarm propagation through a model of the considered system [3], whereas some blind methods rely on machine learning techniques, without system modeling [4].

This paper describes principles and outcomes of a modular approach to fault diagnosis in optical access networks, benefiting from both probabilistic modeling and machine learning. All implementation details of the probabilistic model and of the machine learning algorithm used are given in [5][6][7]. This paper provides a synthesis of main principles, outcomes and insights of this work. The reader can refer to above mentioned papers for all implementation details.

## II. PROBABILISTIC MODELING OF A GPON SYSTEM

Model-based approaches for fault diagnosis rely on a representation of the network structure and its behavior, in particular alarm propagation and correlation [3][8]. Compared to rule-based expert systems, model-based approaches cover more diverse fault configurations than expert systems and can manage unforeseen issues. When the model is probabilized, e.g. with a Bayesian network formalism [9], it gets the ability to deal with uncertainty resulting from non-deterministic fault propagation. It also becomes robust to missing data [7]. This is particularly interesting in network management situations, for which collected monitoring data is often incomplete and depends on the specific network conditions which have led to alarms.

A modular approach for fault diagnosis has been proposed and fully described in [5]. It relies on a 3-layer probabilistic

modeling of the considered system with Bayesian networks. A Bayesian network [9] is a directed acyclic graph the vertices of which are random variables (i.e. alarms, counters, intermediate and root causes in our diagnosis problem) and the edges of which represent conditional dependencies between these random variables. A very simple example of Bayesian network is illustrated in Fig. 1.

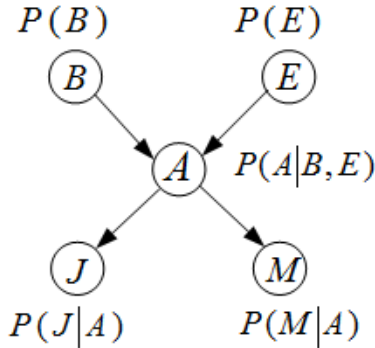


Fig. 1. Example of Bayesian network.

The 3-layer probabilistic model detailed in [5] has been applied to fault diagnosis of a Gigabit capable Passive Optical Networks (GPON) system, composed of Optical Line Termination (OLT) in the central office, Optical Network Terminations (ONT) at customer premises and a (passive) Optical Distribution Network (ODN) in-between, as illustrated in Fig. 2. The model was implemented in Python and called Probabilistic tool for GPON-FTTH Access Network self-Diagnosis (PANDA). PANDA tool does not include Residential Gateways (RG) connected to ONTs, but the modularity of the approach would easily allow this scope extension. Also, a refinement of the model, not addressed in this paper, includes several splitter stages in the ODN. This refinement and related detailed results are presented in [10].

Based on collected monitoring data (e.g. alarms, counters, optical powers), the PANDA model infers missing data and then the most probable root causes of the network fault. The robustness of the model to missing observations is one of its key characteristics compared to a rule-based expert system.

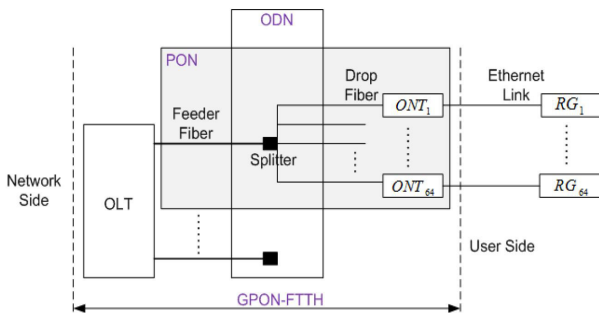


Fig. 2. Structure of a GPON system.

### III. FIRST PANDA IMPLEMENTATION

The structures of the Bayesian networks composing the model and rough estimates of conditional dependencies were

first derived from the GPON standard [11][12] and expert knowledge. This first PANDA implementation was run on a dataset of 10611 diagnosis cases collected from Orange FTTH network, so as to test its alignment with the operational rule-based expert system. The dataset contains alarms typically collected during and after customer calls to the hot line of the access provider.

Table I. gives a confusion matrix crossing diagnosis conclusions obtained with both tools on the 10611 cases [7]. The rows of the table give the numbers of occurrences of diagnosed root causes obtained from the rule-based expert system, whereas the columns give results from the first PANDA implementation.

TABLE I. DIAGNOSIS CONCLUSIONS OF RULE-BASED EXPERT SYSTEM (ROWS) AND FIRST PANDA IMPLEMENTATION (COLUMNS) OVER 10611 CASES

Root causes	1	2	3	4	5	6	7	8	9
1. No default	7210		183	39			17		
2. Faulty ONT		3							
3. ONT configuration mistake			0						
4. Drop fiber attenuated				72			18		
5. Drop fiber broken					1463				
6. ONT power supply failure	2					780			
7. Feeder fiber attenuated							0		
8. Feeder fiber broken		1						57	
9. Unknown root cause	716	4		19		27			0

A first important result is that PANDA approach derives a diagnosis for all tested cases, even if some monitoring data is missing or incomplete. However, it is worth noting that the 716 cases of row 9 and column 1 (i.e. unknown root cause for the rule-based system and no default for PANDA) can be partly explained by the much larger scope of the rule-based expert system, which encompasses equipment beyond the GPON system itself. Nevertheless, the ability of PANDA to deal with uncertainty and missing observations and to manage unforeseen situations is a key property.

Another outcome from table I. is that diagnosis decisions of PANDA and of the rule-based expert system are aligned in more than 97% of all the cases effectively diagnosed by the rule-based engine (the first 8 rows of the confusion matrix). The analysis of the remaining 3% cases (off-diagonal numbers in the confusion matrix, excluding the last row) provides key insights. 183 cases detected as ONT configuration mistakes by PANDA are not managed by the expert system, which has no dedicated rule for these situations. PANDA also diagnoses 39 cases of drop fiber attenuation for which no default is found by the rule-based expert system. In those cases, a low received optical power was measured either at the OLT or at the ONT under study, whereas the received optical power in the reverse

direction was missing. Additionally, in those 39 cases, optical powers related to all ONT neighbors in the PON tree were measured and nominal. In those cases, in spite of one missing optical power measurement, PANDA is capable to derive a diagnosis decision which is fully compliant with the conclusion that would be drawn by a human expert. This capability is allowed by a global analysis of the PON which is performed by the Bayesian inference engine [7].

#### IV. IMPROVEMENTS WITH MACHINE LEARNING

The conditional dependencies of PANDA model have then been finely tuned by machine learning thanks to an Expectation Maximization (EM) algorithm which is detailed in [6]. This EM algorithm was run on 5121 diagnosis cases collected from Orange FTTH network and used as a learning dataset. The principle is to mine the dataset so as to learn the conditional dependencies which will afterwards enable Bayesian inference on other diagnosis cases. Expectation Maximization implements maximum likelihood estimation for an incomplete dataset, and leads to model parameters that "best match" the dataset [6]. The initial model parameters used for running the EM algorithm were those of the first PANDA implementation presented in the previous section.

This second PANDA model, tuned by machine learning, has been assessed on another dataset (test dataset) of 5490 diagnosis cases and compared to the initial PANDA model without machine learning. Table II. gives the confusion matrix crossing diagnosis conclusions obtained with both PANDA versions on the 5490 cases of the test dataset [7].

TABLE II. DIAGNOSIS CONCLUSIONS OF FIRST (ROWS, NO LEARNING) AND SECOND (COLUMNS, WITH LEARNING) PANDA IMPLEMENTATIONS OVER 5490 TEST CASES

Root causes	1	2	3	4	5	6	7	8	9
1. No default	4030			6		7	9		
2. Faulty ONT		0							
3. ONT configuration mistake			183						
4. Drop fiber attenuated				56					
5. Drop fiber broken				14	602		1		
6. ONT power supply failure						402			
7. Feeder fiber attenuated		148					32		
8. Feeder fiber broken								0	
9. Unknown root cause									0

The tuning of the PANDA model through machine learning changes the diagnosis decisions in 3% of the cases, leading to more consistency in some cases, and more clear-cut decisions in other cases. In 14 cases, PANDA tuning changed the decision from "drop fiber broken" to "drop fiber attenuated". In all those cases, a "loss of frame" alarm was observed and no optical power measurement could be collected. Those cases

indeed characterize a low power level which can result in high bit error rate and then a loss of frames: machine learning thus allowed more consistent decisions in those cases. In 148 other cases, the decision is changed from "feeder fiber attenuated" to a more clear-cut "faulty ONT" decision. In those cases, only one ONT is connected to the OLT port, upstream and downstream "loss of signal" alarms are observed, but no optical power measurement could be collected. The "faulty ONT" decision is fully reasonable for a human expert, although those situations could also result from fiber attenuation [6].

#### V. CONCLUSION

A 3-layer probabilistic model based on Bayesian networks was applied to fault diagnosis of GPON systems. With model parameters roughly estimated by skilled humans, very satisfying diagnosis results are obtained on operational GPON FTTH data. The tuning of the model thanks to machine learning leads to even better results in some cases, while some uncertain situations give rise to more clear-cut diagnosis. Outcomes and perspectives of this work are further detailed in [5][6][7] and will be presented at the conference.

#### REFERENCES

- [1] M. Steinder and A. S. Sethi, "A survey of fault localization techniques," in computer networks. Science of Computer Programming, 53:165–194, January 2004.
- [2] L. F. Pau, "Survey of expert systems for fault detection, test generation and maintenance," Expert Systems, 3:100–110, April 1986.
- [3] R. D. Gardner and D. A. Harle, "Alarm correlation and network fault resolution using the Kohonen self-organising map," Global Telecommunications Conference (GLOBECOM 1997), pp. 1398–1402, 1997.
- [4] J. R. A. Goel and P. Sadayappan. Towards a 'neural' architecture for abductive reasoning. IEEE International Conference on Neural Networks, pages 681–688, 1998.
- [5] S. R. Tembo, J. L. Courant and S. Vaton, "A 3-layered self-reconfigurable generic model for self-diagnosis of telecommunication networks," 2015 SAI Intelligent Systems Conference (IntelliSys), London, 2015, pp. 25-34.
- [6] S. R. Tembo, S. Vaton, J. L. Courant and S. Gosselin, "A tutorial on the EM algorithm for Bayesian networks: Application to self-diagnosis of GPON-FTTH networks," 2016 International Wireless Communications and Mobile Computing Conference (IWCMC), Paphos, 2016, pp. 369-376.
- [7] S. R. Tembo, S. Vaton, J. L. Courant, S. Gosselin, M. Beuvelot, "Model-Based Probabilistic Reasoning for Self-Diagnosis of Telecommunication Networks: Application to a GPON-FTTH Access Network," Journal of Network and Systems Management, pp. 1-33, Dec. 2016
- [8] S. Kätker and K. Geihs, "A generic model for fault isolation in integrated management systems," Journal of Network and Systems Management, 5(2):109–130, 1997.
- [9] J. Pearl, "Bayesian networks: A model of self-activated memory for evidential reasoning," in 7th Conference of the Cognitive Science Society, 1985.
- [10] S. R. Tembo, J. L. Courant, S. Vaton and S. Gosselin, "A highly adaptable probabilistic model for self-diagnosis of GPON-FTTH access networks," 2016 24th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, 2016, pp. 1-5.
- [11] Telecommunication Standardization Sector of ITU, G.984.3 Recommendation. ITU-T, 2008.
- [12] Telecommunication Standardization Sector of ITU, G.988 Recommendation. ITU-T, 2012.