ERCIM "Alain Bensoussan" Fellowship Scientific report

Fellow:	Jérôme Lapuyade-Lahorgue
Visited Location:	VTT, Espoo, Finland
Duration of visit:	15/09/2009 - 15/06/2010

Introduction

In this report, I will describe my scientific and research activities during my ERCIM Fellowship at VTT in Espoo (Finland). This fellowship has been spent from September, 15, 2009 to June, 15, 2010, under the direction of Ilkka Norros. The main topic of this fellowship, which will be detailled in this report, relates dependency and reliability of the internet network.

I - Scientific activity

a. Problematic

The functioning of a network is modelled in terms of on-off processes which represent the working state of each component in function of time as illustrated in the figure 1. Using on-off process is looking at the network in a reliability point of view. More particularly, we are interesting in the stationary probability of joint failures and their durations. Finally,

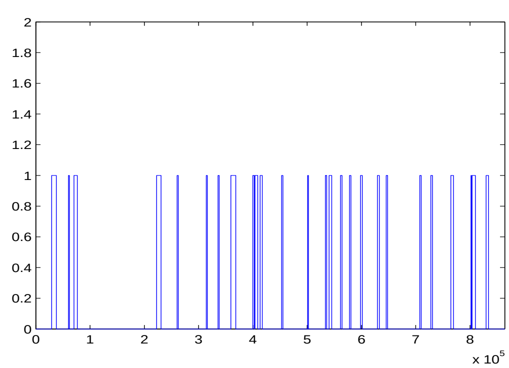

Figure 1: Example of on-off process for one component, the value 1 corresponds to no working of the component and the value 0 to working of the component.

the global network is identified as a multivariate process whose the marginals are the on-off processes of the components. The objective is to propose a good dependent model aiming to represent the dependencies between such processes. Indeed, the dependencies between such processes are motivated by the fact that if one component is failed, it can influence the other components and increase their probability of failure.

b. State of art

The model that I experimented during this fellowship extends a previously studied model in which components of the network are independent. This model is defined as follows. Suppose that the network is composed of K components, which are represented as K on-off processes $(X_t^{(k)})_{t\in\mathbb{R}}$ taking their values each in $\{0,1\}$, where 0 represents the on-state of the component and 1 the off-state. The number of failed components of the system at time t is then given by $Y_t = \sum_{k=1}^{K} X_t^{(k)}$. Each on-off process is a stochastic process defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and we model each independent component as an exponential-Pareto distributed duration system. It means that the sojourn duration of the state 0 is an exponential distribution with parameter λ_j for the component j and the sojourn duration of the state 1 is a Pareto distribution whose parameters are $k > 0$ (minimum duration) and $\alpha > 1$ (shape parameter). The study of the joint process which represents the network can be successfully achieved in this independent case using the Palm calculus. The Palm calculus is an elegant mean to get stationary distribution of continuous time jump processes. The Palm theory consists in defining the model under a probability called Palm probability and computing the stationary distribution under the probability $\mathbb{P}$ using a formula called "inversion formula". For the independent components model, the problem is easy to solve. Indeed, one can compute stationary distributions for each individual component isolated from the system, and the stationary distribution is a product of the individual stationary distributions. Moreover, in the independent case, the joint Palm probability can be expressed easily from the individual Palm probabilities, which allows us to compute the sojourn durations for all failure configurations. However, things are not so easy in the dependent model that we have defined, and some approximations and other theories are required to solve it.

c. Dependencies between components

The dependent model that I dealt along my fellowship is a two components system. The "off" state sojourn duration for both components follows the same Pareto distribution, whereas the "on" state sojourn duration for component l follows an exponential distribution of parameter λ'_l if the other component is working and of parameter $\lambda_l > \lambda'_l$ if the other component doesn't work. The condition $\lambda_l > \lambda'_l$ decreases the "on" state sojourn time if the other component fails. Despite the apparent simplicity of this model, Palm calculus is not enough to solve it. Indeed, let $\mathcal{P}_l$ be the Palm probability for the component l and $\mathcal{P}$ be the Palm probability for the joint process. If we denote $\lambda^{(l)}$ the mean number of changes of state in the interval $(0,1]$ for the component l and $\tau_{i,j}$ the mean of sojourn duration for the joint state (i,j), the stationary distribution of the joint state (i,j) is given by:

$$\mathbb{P}(X_t = (i,j)) = \left[\lambda^{(1)}\mathcal{P}_1(X_0 = (i,j)) + \lambda^{(2)}\mathcal{P}_2(X_0 = (i,j))\right]\tau_{i,j}, \tag{1}$$

and the difficulty is implied by the difficulty of computation of $\tau_{i,j}$. This quantity is difficult to compute due to the long memory behaviour implied by the simultaneous use of Pareto distributions and dependency between the marginals.
In order to overcome this difficulty, I have proposed different approximations of this model. For the first models that I proposed, I established that a two-component model is formed of independent marginals under two conditions: a condition on the sojourn duration and a condition on the embedded process. Consequently, one can build a dependent two-component model by withdrawing one of the two conditions. In these two models, the stationary distributions and sojourn durations can be computed using Palm calculus. We have validated these models by comparing theoretical values of the stationary distributions and sojourn durations and their experimental values where the data have been simulated according to the previous model. Unfortunately, the obtained results were poor, consequently we have proposed another approximation. In the second approximation, we consider that the state $(1, 1)$ which corresponds to a joint failure, transits directly to the state $(0, 0)$. This direct transition is interpreted as a simultaneous recovery of the working for both components. This model appears to be a kind of semi-Markov model and the stationary distribution and the sojourn durations can be computed without using Palm theory but ergodic theory of semi-Markov chains. The experiments confirm better results than for the first approximations. Finally, in the last part of this fellowship, I have interpreted the discrete time version of the first studied general model as a marginal of a Markov chain by introducing two auxiliary processes representing the remaining sojourn time in the same state for each components. Despite of the interest of this method, the stationary distribution of the resulting Markov chain is heavy to compute by solving Kolmogorov equations, due to the fact that this Markov chain has a countable but infinite number of atoms.

d. Conclusion and perspectives

This first task constitutes a step to future research. We have seen that despite of the apparent simplicity of the model, the solution of it is not so simple to compute and requires approximations and simplifications of the model. This study can be used as a starting point by future fellows and researchers, also for proposing other realistic dependent models. Moreover the extension in the case of several components has to be proposed. A system with multiple dependent components can be represented as a graph of dependencies whose state evolves according to the time. Each neighbour of a node influences the node itself as in an epidemic propagation system.

II - Publications during fellowship

Two papers on preparation:

1. **Paper 1** (to a journal): It concerns more or less existing VTT work: Palm approach to on/off process, independent case, definition of Poisson-Pareto dependent models with many components, the special case of two components with "component-independent stress factor", and an associated inequality relating this to the general two-component cases;

2. **Paper 2** (to a conference): It focuses only on dependent case with the same model as in paper 1. Description of the model as a Markov process, simulations of many component cases, simplifications and approximations of the model and numerical comparisons.