Genetic algorithms for delays evaluation in networked automation systems
Boussad Addad, Saïd Amari, Jean-Jacques Lesage

To cite this version:
Genetic algorithms for delays evaluation in networked automation systems
B. Addad *, S. Amari, J-J. Lesage

Abstract
In this paper, we present an approach to evaluate end-to-end delays in packets switching networked automation systems. Since Client-Server paradigm is considered for communication between the field devices, the existing methods of network delays evaluation are hardly applicable to assess realistic upper bounds of these delays. In an effort to enhance these delays evaluation, we propose an alternative method. Two algorithms, usually used for optimization problems, exhaustive and genetic algorithms, are then developed to achieve this purpose. While a formal proof about the capacity of the former one to ensure the worst delay overestimation is given, the latter proves to provide faster and more accurate results at the same time. This is shown on a practical case study while comparing the results of the two methods.

Keywords: Genetic algorithms, Delays evaluation, Networked automation systems, Switched packets, Client-Server protocol.

1. Introduction

Fieldbus systems have been successfully introduced in industrial automation, ensuring real time requirements on one hand and devices safety on the other hand. Nowadays, the trend is also to use the same communication technology at different levels in the industrial organizations: management and automation. A solution that supports such a vertical integration has to be able to provide high throughputs in the upper level as well as small and accurate response times in the field level. The Ethernet solutions, which were initially developed to office networks, can be considered as such a new generation of fieldbuses. Currently, many automation producers and alliances developed their own industrial Ethernet standard (Neumann, 2007). Each solution with a specific protocol is best suited to a particular application. A Client/Server protocol like Modbus over Ethernet, even not adequate for strict real time applications like motion control, is a simple and a reasonable solution for many purposes in industrial control systems. Indeed, it is an application protocol (the 7th layer of the OSI model) that is completely compliant with the standard Ethernet. Therefore, vertical integration is easily achieved. Thus, high level functions like diagnosis and device management are easily implemented. Unfortunately, with such a protocol, no global medium access scheduling is available and different delays due to waiting for resources availability or synchronization are caused. So, the evaluation of its time performances like the response time is complex and the investigations that deal with this problem are rare. The existing methods are often studies of particular systems based on model checking (Greifeneder and Frey, 2007; Witsch et al., 2006; Ruel et al., 2009) that suffers from the classical state explosion problem. Another method is based on high level colored Petri nets simulation (Marsal et al., 2006; Zaitsev, 2004). This method is time driven and very onerous of time. Moreover, it does not provide a formal analysis or proof about its capacity to sweep the worst scenarios corresponding to the worst delays. Finally, an experimental method using a logical network analyzer dedicated to delays measuring is presented in Denis et al. (2007). Hence, the aim of the current paper is to propose an adequate method to assess upper bounds of end-to-end delays of switched packets in the context of Client-Sever automation systems. While a proof about the capacity of the method to assess the worst delays is provided, a genetic algorithm is developed to achieve it much faster.

The remainder of this paper is organized as follows. Section 2 introduces the context of our investigation and the motivations to develop a method for end-to-end delays evaluation. Thereafter, two algorithms, to look for the worst delays, are developed in Section 3: an exhaustive algorithm in Section 3.1 and a genetic algorithm in Section 3.2. A case study is then considered to perform a comparison between them in Section 4. Finally, Section 5 addresses some concluding remarks.

2. Client server automation systems over switched Ethernet

The studied automation architecture works according to Client/Server protocol. It is constituted mainly of PLCs (programmable logic controllers), RIOMs (remote input output modules) and a switched Ethernet network that enables communication between...
all the components (Fig. 1). The PLCs (clients) send periodically requests to the RIO Ms (servers) and wait for answers. When a RIOM receives a request, it puts it in a queue so as to process all the waiting requests according to a FIFO policy. A request is mainly to get information from the plant (e.g. is the maximal level of water reached?) or to provide orders (e.g. close the valve) or a combination of both. A major criterion of time performance of such a system is the response time $T_r$ (Fig. 1). It is defined as the delay between the occurrence of an event in the plant (e.g. the maximum level is reached) and the arrival of the consequence, issued from a controller, on the controlled plant (closure of the valve). The response time can be defined intuitively as the reactivity delay of the automation system.

As a matter of fact, the evaluation of the response time of these systems is tricky. Indeed, different delays due to non synchronization of the components, resources sharing and of course the intrinsic delays (processing) are to be considered.

In Addad et al. (2010), we developed an analytic method to evaluate this response time. A formula giving an upper bound of the response time is obtained. Obviously, the delays we call the upper bounds of the end-to-end delays. Indeed, a small overestimation of the delays may lead to a huge overestimation of the response time.

A typical end-to-end delay is the time for a request to cross the switches from its generation by a PLC until its arrival to a RIOM. Therefore, to assess an upper bound of the response time, upper bounds of these end-to-end delays are needed. It could be thought then that existing methods like the well known network calculus (Cruz, 1991; Le Boudec and Thiran, 2004; Georges et al., 2005) or worst case methods (Fan et al., 2008; Lee and Lee, 2002) can be used for this purpose. Unfortunately, in the context of switched packets, combined to Client-Server paradigm, it is not so obvious. Indeed, the flows independency supposed by the previous methods is not verified. For instance, a RIOM does only answer a received request and therefore a request with its corresponding answer cannot exist at the same time in the system.

Besides this impeding fact, the formula is very sensitive to the upper bounds of the end-to-end delays. Indeed, a small overestimation of these delays may lead to a huge overestimation of the response time upper bound. As a result, the quality of control in these automation systems are dramatically degraded since the control law synthesis is based mostly on the upper bound assessment (Addad and Amari, 2008). Thereby, an adequate method to evaluate the end-to-end delays in such systems is to be investigated. This is the objective of the next section.

3. End-to-end delays evaluation

As aforementioned, the studied automation architecture works according to Client-Server protocol. The PLCs, which are the clients, send requests and the RIO Ms (the servers) return answers accordingly. To explain the proposed method of delays evaluation, we use the system in Fig. 1.

- PLC1 sends a burst of three requests (series of three frames sent one after another) periodically with a period $T_1$ to the modules R12, R13 and R14 (Figs. 1 and 2).
- PLC2 sends a burst of three requests periodically with a period $T_2$ to the modules R14, R15 and R16.
- PLC3 sends a burst of two requests periodically with a period $T_3$ to the modules R16 and R17.

Actually, the PLCs which are the sources of frames (we use the term of "senders") are neither synchronized nor scheduled and therefore can start sending their bursts at any time. A scenario is shown in Fig. 2 where lags $\tau_2$ and $\tau_1$ with respect to the start-sending date of PLC1 (chosen as a reference) are represented. Obviously, different lags (different scenarios) will lead to different end-to-end delays within the system. Therefore, the main issue is then to find the critical scenario (the lags $\tau_j$) that causes the maximum delay of a given frame. This is the purpose of the following algorithms.

3.1. Exhaustive exploration algorithm

As explained above, the interference between the bursts is entirely up to the lags $\tau_j$ between the senders. Intuitively, a sender will influence another if it starts sending frames at more or less the same time. A given frame experiences a delay depending on the set of frames that preceded it in a switch and are waiting to be forwarded first. In the context of our study where the pattern of generation of the frames from PLCs is known beforehand (periodically), all the possibilities of this set of frames can be identified exhaustively. In Fig. 3 for example, the number of possibilities of the set $E$ of the frames of parallel bursts that enter before a frame $f^*$ is equal to $4 \times 3 = 12$ (4 frames in the first burst and 3 in the second one). It is a simple combinatorial operation. As we notice in Fig. 3, there are infinite situations that correspond to the same set $E$, but obviously not the same delays. Therefore, the issue is to handle the variation of the delays from a scenario to another. The next lemma and theorem will show that the discrepancy from a scenario to another, corresponding to the same set $E$, is at worst equal to a well determined length.

**Lemma.** Suppose $f$ a frame that enters a switch at time $t^*$ and experiences a delay $\Delta$ in it. Let $f'$ be a frame (of a parallel burst) that comes immediately before $f$ and enters the switch at time $t^* - \delta_r$.

![Reference time](image)

**Lemma.** Suppose $f$ a frame that enters a switch at time $t^*$ and experiences a delay $\Delta$ in it. Let $f'$ be a frame (of a parallel burst) that comes immediately before $f$ and enters the switch at time $t^* - \delta_r$.

![Interference domain 1](image)

![Interference domain 2](image)

![Sending requests](image)

![Receiving answers](image)
Theorem. Suppose we look for the worst scenario by searching the lags $\tau_i$, in an enough large domain.

If a step $\delta_i < \delta$ is used to vary the lags of the start-sending dates $\tau_i$, in a combinatorial scheme, then an exhaustive exploration is achieved and the accuracy of the result is worth $\delta_e$. In other words, the gap between the assessed delay and the effective worst one is smaller than $\delta_e$.

Proof. From the lemma, the maximal delay is reached if $f^*$ enters the switch a bit after $f_i$. Since the step $\delta_i$ is smaller than the inter-arrival time, then we are sure to get the set $E$ and the best case is when $f^*$ enters $\delta_i$ before $f_i$. Hence, if every frame $f^*$ arrives $\delta_i$ before $t^*$, then the processing of the set $E$ will be finished, in the best case, $\delta_e$ earlier. Thus, the assessed delay added up to $\delta_e$ is at worst $\delta_e$ greater than the real delay.

Remark 3.1. With Ethernet $\delta = \min_n((72+12)/C_k)$ where 72 bytes is the minimal frame length (preamble included), 12 bytes is the inter-frames gap (96 bits or 12 bytes) and $C_k$ the physical capacity of the port $k$ of a switch.

Thereby, we can use a step $\delta_e$ respecting the previous theorem condition, to vary the lags $\tau_i$ in a domain noted $[-T_{Domain}, +T_{Domain}]$ where $T_{Domain}$ is taken large enough to include the worst scenario (assessed for instance using a pessimistic method).

Hence, for each combination of the lags, we simulate the behaviour of the system and assess the corresponding delay $D_{end2end}$ (see Fig. 4). This is represented by the block $Simulator(\tau_1, \tau_2, ..., \tau_n)$ in Fig. 4 with the lags $\tau_i$ being introduced as entries. For a system with $n$ senders (PLCs), the diagram of the exhaustive exploration method is drawn in Fig. 4. The maximum of all the obtained delays is then chosen as an upper bound. The simulation can be achieved using any suitable simulation method. In the case of the current study, a virtual queuing based simulator, already applied in Addad and Amari (2009), has been used for this purpose.

It has the advantage of being an event driven simulator and therefore much faster than the usual time driven simulators. Despite this advantage, the simulation may last a long time when dealing with very large systems. Indeed, the number of times to run the block $Simulator(\tau_1, \tau_2, ..., \tau_n)$ is exactly equal to $(2 \cdot \lceil T_{Domain}/\delta_e \rceil)^n$ where $n$ is the number of senders and $\lceil T_{Domain}/\delta_e \rceil$ the integer part of $T_{Domain}/\delta_e$. Thus, this number grows exponentially when the number of senders grows linearly. While this exhaustive method is quite satisfactory in some cases, it can be very onerous of time in others, so other more efficient methods are to be investigated.

Genetic algorithms are then considered as an alternative in our study.

3.2. Genetic algorithms

Genetic algorithms (GAs) were first introduced by Holland (1975) and have been extensively explored later in other
investigations for optimization problems resolution in almost all domains. With regard to fieldbus systems, GAs were used in Georges et al. (2006), Zhang and Zhang (2007) and Carro-Calvo et al. (2010) to optimize the partition of switched Ethernet networks into sub-networks where the intra sub-network communication traffic is maximized whereas the inter sub-networks communication is minimized. They were used in Lee et al. (2004) to help the designer to select the optimal timers in a Proflibus Token Passing Protocol fieldbus so as to satisfy the maximum allowable delays of the real time data on one hand and maximize the non real time data transfer on the other hand. They were also used in Cheng and Yang (2010) and García-Nieto et al. (2010) to investigate the problem of dynamic quality of service in mobile ad hoc networks. We can also quote (Saniee Abadeh et al., 2007) where GAs were applied GA in our problem:

A GA is a structured stochastic optimum search method that mimics the process of biological evolution. At the beginning, a GA is initialized with a collection of sets of parameters. Each set is called a chromosome whereas a parameter a gene. Then, the chromosomes are evaluated according to their fitness of solving an optimization problem (maximization of \( D_{end2end} \) in our case). At every generation (algorithm execution time), the fittest chromosomes (parents) are allowed to mate (crossover) and bear offspring (children). Then, the best children become the parents of the next generation. The previous steps are repeated until finding a satisfactory solution of the problem. To avoid falling in a local optimum as is often the case with the traditional hill-climbing search methods, a mutation mechanism is added. Depending on the optimization problem, many ways exist to accomplish each step of the GA: codification of the chromosomes, initialization of the population, selection of the parents, crossover and mutation. Here are the features of the applied GA in our problem:

- **Chromosomes encoding**: The parameters handled in our problem are the start-send dates or the lags \( \tau \) which are real numbers. Therefore, we apply a continuous GA and the genes are simply real numbers. The \( k \)th gene noted \( g_k \) of each chromosome \( i \) represents the lag \( \tau_k \). The \( i \)th chromosome is noted \( \Theta^k = (g_k^1, g_k^2, \ldots, g_k^3) \) and its fitness \( f(\Theta^k) \) where \( f(\Theta^k) = D_{end2end} \).
- **Initialization of the population**: In some automation systems, we may have an idea about the worst scenario and therefore initialize the GA with the corresponding parameters. In complex systems, it is rarely the case. Therefore, a random initialization of the parameters in the domain \([-T_{domain}+T_{domain}] \) is considered in the current study.
- **Selection of the parents**: According to natural selection theory, the fittest members with the greatest fitness survive and the others die. One way to represent this strategy is to use the so-called roulette-wheel-spinning. A unit circumference is cut into slices, each slice being dedicated to one member according to its relative fitness. The \( i \)th member is then associated to the portion \( P_i = f(\Theta^i)/\sum_k f(\Theta^k) \). Thereafter, we spin the wheel and if the pointer stops at the \( j \)th portion, the chromosome \( j \) is selected and placed into a mating pool. Obviously, the fittest members with the biggest portions are more likely to be chosen than the others.
- **Crossover (mating)**: Pairs are chosen from the mating pool and mate with a given probability \( P_{cs} \) to form the offspring. Many approaches exist to deal with the crossover in continuous GA (Passino, 2005). One of the most efficient and simplest methods is combining two genes \( g^1_k \) and \( g^2_k \) of the parents \( \Theta^1 \) and \( \Theta^2 \) to form the children \( Ch^1 \) and \( Ch^2 \) (with genes noted \( Ch^1_k \) and \( Ch^2_k \)) as follows:

\[
\begin{align*}
Ch^1_k &= 2g^1_k - (1 - \alpha)g^2_k \\
Ch^2_k &= 2g^2_k + (1 - \alpha)g^1_k
\end{align*}
\]

\( \alpha \) being a random real number in the domain \([0,1]\).

- **Mutation**: To avoid a quick convergence of the GA and the risk of being trapped in a local optimum, we force it to explore other regions of the space by introducing random changes or mutations in the genes. The process of mutation on a gene \( g_k^i \) of a member \( \Theta^i \) can be performed with a probability \( P_{mut} \) as follows:

\[
g^i_k = T_{domain}(2\beta - 1)
\]

\( \beta \) being a random real number in the domain \([0,1]\). The mutation is applied on all the members of the population except for the best member of the generation if elitism option is chosen. With elitism option, the best member of a generation is conserved for the next generation without any change. As a consequence, the fitness function is monotonic along the generations (non-decreasing function as in Fig. 7).

The process described previously is repeated until a satisfactory solution is found or the best fitness does not change during a minimum number of generations. The GA applied for our problem resolution is implemented using Matlab software and the main steps are shown in Fig. 5. The following notations are adopted:

- **PopNum** is the total number of members of the population.
- **MaxGen** is the maximal number of generations.
- **PCss** is the probability of crossover.
- **PMt** is the probability of mutation.
- **rand** is a random number from the domain \([0,1]\).

Remark 3.2. As can be seen in Fig. 5, the same block representing the simulator used with the exhaustive method is used with GA. Indeed, this block is only used to simulate the behaviour of the system for a given scenario so as to assess the corresponding delay. Note that using a simulator and introducing manually the entries (the lags \( \tau_k \) of each scenario would not be viable given the number of times the block is to be run. In this study, both the simulation block and optimization methods (either exhaustive or GA) are implemented using the same programming language (Matlab) and the whole process is therefore automated.

### 4. Case study

The system of the case study is shown in Fig. 6. It works as follows: three controllers, PLC1, PLC2 and PLC3 scan, respectively, the ordered sets of RIOMs: \( \{R_{12}, R_{13}, R_{14}\}, \{R_{12}, R_{13}, R_{16}\} \) and \( \{R_{16}, R_{17}\} \) with 72 bytes requests. This is the automation part we used for explanation in Section 2, but as stated previously, the studied system presents the advantage of ease of high level functions integration since standard Ethernet is considered for communications. Therefore, a second part \( PART \ II \) is added for this purpose; PC1 exchanges long frames of 1008 bytes length with PC3 and PC4 whereas PC2 has the role of a supervisor of the first part \( PART \ I \). It scans all the RIOMs every 300 ms with 72 bytes requests so as to check the good functioning of the modules. The aim is to evaluate an upper bound of the delay \( D_{end2end} \) that a request, sent from PLC2 to R14, experiences in the network (Fig. 6). Any request would be chosen, but we selected this one since R14 is shared by PC2, PLC1 and PLC2. Thus, particular and not trivial scenarios lead to the worst delay. Therefore, we should find them using different methods. To evaluate the delay in question, we applied the two previously
exposed methods, exhaustive exploration and GA. We also considered two cases using the exhaustive method: Exhaustive I with step $\delta_e$ being set up to 20 ms at the beginning and to 1 ms subsequently whereas in Exhaustive II this step is progressively narrowed from 50 to 1 ms. The results of assessment using the different approaches are reported in Table 1.

As we can note, all the methods lead to nearly the same assessment of the worst end-to-end delay. The duration of simulation is however very different. First, the use of a small step in the exhaustive exploration may lead to long simulation durations (Exhaustive I). Indeed, the smaller the step $\delta_e$ is, the greater the number of situations to check is (recall that this number is equal to $2^{2\Delta T_{Domain}/\delta_e}$).

A more interesting method is to begin with a relatively large step $\delta_e$ (50 instead of 20) if possible (it must check the condition given by the theorem) and narrow it progressively. This is made in Exhaustive II and we notice that the simulation duration is dramatically reduced from ~2 h to about 4 min. We can also note that the gap of underestimation in all the cases is smaller than the step $\delta_e$. Indeed, the sum of the used step and the assessed delay is always greater than the final assessed delay. In Exhaustive II for instance, we have an assessed delay of 283.52 ms with a step $\delta_e$ equal to 50 ms. The sum of both is greater than the final delay, which is equal to 319.96 ms. This feature can be checked in all the results of Table 1. Naturally, this corroborates the statements of the theorem of Section 3.1.

On the other hand, the genetic algorithm not only provides an accurate upper bound of the delay (see Fig. 7), but also achieves it in relatively negligible simulation duration of about 29 s. Moreover, a gap of 1% from the final maximal delay is reached in almost 100

---

**Fig. 5. GA diagram.**

---

**Fig. 6. Case study: evaluation of delay $D_{\text{max}}$.**

---

**Fig. 7. The best fitness (maximal end-to-end delay) along the generations.**
5. Conclusion

In this paper, we proposed an approach to evaluate end-to-end delays of switched packets in networked automation systems. The Client-Server paradigm being used as the protocol of communication, the existing methods of delays evaluation are indeed hardly applicable. So, an exhaustive exploration method and genetic algorithms were developed to assess upper bounds of these delays. On a case study, we showed that both methods give almost the same results but a clear advantage of using GAs has been noticed since the simulation durations are dramatically different. While the exhaustive one may be preferred when dealing with simple automation systems, given its simplicity of implementation. GAs are more suitable for evaluation when large architectures are under consideration.

References