Quantifying the Similarity of Non-bisimilar Labelled Transition Systems

Gwen Salaün

Univ. Grenoble Alpes, CNRS, Grenoble INP, Inria, LIG, F-38000 Grenoble France

Abstract

Equivalence checking is an established technique for automatically verifying that two behavioural models (Labelled Transition Systems, LTSs) are equivalent from the point of view of an external observer. When these models are not equivalent, the checker returns a Boolean result with a counterexample, which is a sequence of actions leading to a state where the equivalence relation is not satisfied. However, this counterexample does not give any indication of how far the two LTSs are one from another. One can wonder whether they are almost identical or totally different, which is quite different from a design or debugging point of view. In this paper, we present an approach for measuring the similarity between two LTS models. The set of metrics is computed automatically using a tool we implemented. Beyond presenting the foundations of the proposed solution, we will show how it can be applied to two concrete application domains for supporting the construction of IoT applications on the one hand and for contributing to the process model matching problem on the other.

Key words: Equivalence Checking, Labelled Transition Systems, Quantitative Analysis, Similarity Measure, Internet of Things, Process Matching.

1. Introduction

Designing and developing distributed software has always been a tedious and error-prone task, and the ever increasing software complexity is making matters even worse. Although we are still far from proposing techniques and tools avoiding the existence of bugs in a software under development, we

know how to automatically chase and find bugs that would be very difficult, if not impossible, to detect manually.

Model checking [1] is an established technique for automatically verifying that a model (Labelled Transition System, LTS), obtained from higher-level specification languages such as process algebra, satisfies a given temporal property. Equivalence checking [16] is an alternative solution to model checking and is very helpful to check that two models (requirements and implementation for instance) are equivalent from the point of view of an external observer. When these models are not equivalent, the checker returns a Boolean result with a counterexample, which is a sequence of actions leading to a state where the equivalence relation is not satisfied. However, this counterexample does not give any indication of how far the two LTSs are one from another. One can wonder whether they are almost identical or totally different, which is quite different from a design or debugging point of view.

In this paper, we propose a set of metrics for quantifying the similarity of two behavioural models described using LTS. More precisely, our solution takes as input two LTS models and applies first the partition refinement algorithm [21, 8] to identify bisimilar and non-bisimilar states between the two LTSs. Then, we focus on non-bisimilar states and compute a set of global and local metrics for each couple of non-bisimilar states. This allows us to build a matrix with a measure between 0 (totally different states) and 1 (bisimilar states) for each couple. To do so, we rely on several criteria such as the matching of incoming/outgoing transitions, the similarity of neighbour states, the shortest distance from the initial state, and the distance to the closest bisimilar state. Once this matrix is computed, we use it to finally obtain a global measure of similarity of both LTSs. All these measures are computed automatically using a tool we implemented in Python and applied on a large set of examples.

Better understanding and measuring the difference between two behavioural models can be of interest in many different contexts and application areas. It can be used for debugging purposes when the counterexample is not sufficient for detecting the source of the bug, for measuring the distance between two versions of a software, for process model matching in the context of business process and management, etc. We will show in this paper how it can be helpful in two different areas: Internet of Things and process model matching. One of the main challenges in the Internet of Things (IoT) is to build a new application by composing existing objects or devices. This application or composition is satisfactory if it conforms to what the user ex-

pects from it. These requirements are formalised using an abstract goal in this work. We will show how we use the proposed measures to compare the candidate composition and the goal. The quantitative results help in understanding what parts of the composition are correct or not with respect to this goal, and in guiding the user to finally end up with a satisfactory composition. As far as process matching is concerned, the goal is to compare two processes modelled using the BPMN [7] standard notation. This comparison aims at identifying correspondences and differences between these two models. This comparison is useful in order to support the evolution process when updating one process with another one, and to know exactly what differs between those two before deciding to effectively change one process for the other. We will see with this second application how our approach can help for both detecting similar and different parts in the two process models.

An early version of this paper has been published in [26] and has been significantly extended here as follows:

- a notion of similarity of path has been added as global criteria to obtain better results;
- several global measures have been implemented;
- a second use case shows the application of our techniques to the process model matching domain;
- extended experimental results have been carried out for validation of our approach; and
- the paper has been revised in many places to take the aforementioned extensions and improvements into account.

The rest of this paper is organized as follows. Section 2 defines LTSs and the notion of strong bisimulation used in equivalence checking. Section 3 presents the details of our approach to compute both the similarity matrix and the global measure of similarity. Section 4 illustrates the proposed solution on two case studies, one from the IoT domain and one from the process modelling area. Section 5 reviews related work and Section 6 concludes the paper.

2. Labelled Transition Systems

In this work, we rely on *Labelled Transition System (LTS)* as low-level behavioural model of concurrent programs. An LTS consists of states and labelled transitions connecting these states.

Definition 1 (LTS)**.** *An LTS is a tuple $M = (S, s^0, \Sigma, T)$ where S is a finite set of state identifiers; $s^0 \in S$ is the initial state identifier; Σ is a finite set of labels; $T \subseteq S \times \Sigma \times S$ is a finite set of transitions.*

A transition is represented as $s \xrightarrow{l} s' \in T$, where $l \in \Sigma$. An LTS can be produced from a higher-level specification of the system described using process algebra for instance. Process algebraic specifications can then be compiled into an LTS using specific compilers. We support nondeterministic LTSs in this work, that is, there may be several transitions outgoing from a specific state labelled with the same action.

When comparing two LTSs, we can use different notions of equivalence, from weak ones such as trace or observational equivalence to stronger ones such as strong bisimulation. In this work, we chose to use strong bisimulation as originally defined in [16]. Supporting weaker notions of bisimulations where silent actions are handled separately is part of future work. It is worth noting that both LTSs are reduced using existing minimization techniques before comparing them. As we focus on strong bisimulation in this work, LTSs are minimized modulo strong bisimulation using CADP tools [6] (bcg_min).

Definition 2 (Strong Bisimulation)**.** *A relation R is a strong bisimulation between states in S iff for all $s_1, s_2 \in S$ such that $R(s_1, s_2)$, both conditions hold:*

- $(\forall b \in \Sigma, t_1 \in S)\ (s_1, b, t_1) \in T \implies (\exists t_2 \in S)\ (s_2, b, t_2) \in T \wedge R(t_1, t_2)$
- $(\forall b \in \Sigma, t_2 \in S)\ (s_2, b, t_2) \in T \implies (\exists t_1 \in S)\ (s_1, b, t_1) \in T \wedge R(t_1, t_2)$

Two states s_1 and s_2 are strongly bisimilar (written $s_1 \approx_s s_2$) iff there exists a strong bisimulation R such that $R(s_1, s_2)$. Two LTS $M_1 = (S_1, s_1^0, \Sigma_1, T_1)$ and $M_2 = (S_2, s_2^0, \Sigma_2, T_2)$ are strongly bisimilar (written $M_1 \approx_s M_2$) iff $s_1^0 \approx_s s_2^0$.

Equivalence checking is usually checked using partition refinement algorithms [21, 8]. These algorithms aim at building the minimal number of blocks, where a block is a set of (strongly) bisimilar states. One block is

called an equivalence class. In order to check whether two LTSs are equivalent, the partition refinement algorithm is called with the union of both LTSs as input. At the end of this computation, if both initial states are in the same block, the LTSs are equivalent.

3. Comparing Non-bisimilar States

In this section, we present the measure of similarity between two LTSs. When comparing both LTSs, if the equivalence checker returns *true*, it means that the two LTSs are the same with respect to strong bisimulation. If the equivalence checker indicates that both LTSs are not strongly bisimilar, our similarity comes into play to compute further information. More precisely, in that case, we apply our approach to quantify the difference between the two subparts of both LTSs that are not equivalent. The measure relies on two kinds of criteria, namely global and local criteria, which focus on two non-bisimilar states (one in each LTS). We also present global measures that give a measure of how far both LTSs are. Finally, we introduce a tool that allows us to automatically compute all these results.

3.1. Overall Approach

Given two LTSs, we first use the partition refinement algorithm mentioned in Section 2 to compute bisimilar and non-bisimilar states. Then, we focus on non-bisimilar states and propose a measure comparing all non-bisimilar states according to several global and local criteria. These global and local criteria exploit the different elements present in both LTSs, namely, initial states, labels, structure of the LTS or position of states in the LTS. For each couple of non-bisimular states (one non-bisimilar state from each LTS), we compute a degree of similarity which belongs to [0..1]. All these results are stored in a matrix where non-bisimilar states of one LTS appear in row and non-bisimilar states of the other LTS appear in column.

Given two non-bisimilar states s_1, s_2 where $s_1 \in LTS_1$ and $s_2 \in LTS_2$, we compute the similarity of those states using global and local criteria. Global criteria aim at considering the structure of both LTSs and looking at the respective positions of both states in their LTSs. More precisely, there are three global criteria. The first one computes the distance from the initial state to the given state in both LTSs and compares those distances. The second one computes the distance from a given state to the closest bisimilar state and compares those distances. In both cases, we compute the shortest

distance. The third one aims at comparing whether those paths (from the initial state to the current state or from the current state to the closest bisimilar state) are similar. To do so we first search for the shortest paths in both LTSs and we compute the number of common labels out of the total number of labels, which gives the main part of this similarity measure. We also add a bonus to this resulting value if the order of labels is the same in both paths.

There are four local criteria. The first one compares outgoing transitions to see the number of matching labels. The second does the same with incoming transitions. The third one checks whether the nature of states differ (initial or not). The last one compares the similarity of neighbour states.

Given all these values for a couple of states, we can then compute its value in the matrix ($matrix[s_1, s_2]$). This is obtained by using the weighted average of these values (*e.g.*, 1/7 or arbitrary weights). Note that these weights are parameters of our approach. We can decide to change them for putting more emphasis on labels or on some of the local criteria for instance. In the results and experiments presented in the paper, we chose equal weights for these parameters, because, in the context of our application areas (IoT and process matching), there was no clear argument to emphasis some of the values with respect to the other ones.

Since the similarity of neighbour states uses the matrix itself, we use an iterative algorithm that stops when the matrix stabilizes. Once the matrix is computed, we can compute several global measures, which give a degree of similarity of both LTSs.

In the rest of this section, we explain in more details the metrics used in this work for computing the similarity measure.

3.2. Global Criteria

The two first global criteria aim at comparing two states $s_1 \in LTS_1$ and $s_2 \in LTS_2$ by looking at their positions in their respective LTSs. We rely on two measures: (i) comparison of distance from initial states to states s_1 and s_2 (d_{init}), (ii) comparison of distance between s_1 and s_2 to their closest bisimilar states (d_{bis}). In both cases, we search for the shortest path, which is enough to have an estimation of the position in the LTS while avoiding to analyse all possible paths outgoing of (leading to, respectively) these states. Both measures are then computed in the same way as follows: $1 - (abs(d_1 - d_2)/max(d_1, d_2))$, where d_1, d_2 is the distance from s_1, s_2 to

the closest bisimilar state or from initial states to s_1, s_2. Function *abs* is the absolute value function and *max* returns the longest distance.

Example. We illustrate with a simple example where we take the shortest distance from the initial states to two states $s_1 \in LTS_1$ and $s_2 \in LTS_2$ (Figure 1). Assume first that $d_1 = 1$ and $d_2 = 8$. In that case $d_{init} = 1 - ((8-1)/8) = 0.125$ corresponding to a quite low value for this distance criterion. Consider now that $d_1 = 3$ and $d_2 = 5$. This results in a highest value $d_{init} = 1 - ((5-3)/5) = 0.6$. If we take equal values such as $d_1 = 4$ and $d_2 = 4$, we obtain the highest value $d_{init} = 1 - ((4-4)/4) = 1$, which means that these two states are not distinguishable with respect to this metric.

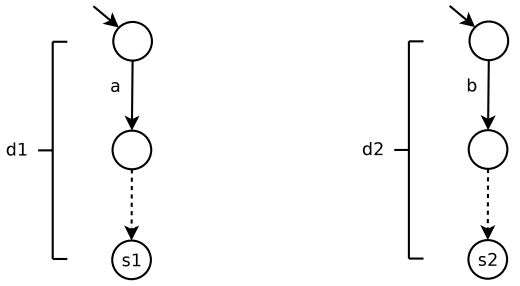

Figure 1: Example for the Distance Comparison Criterion.

The third global criterion aims at measuring the similarity of two paths. Those two paths are taken in both LTSs either from the initial state to the current state or from the current state to the closest bisimilar state. To compute this measure, we first extract from both paths all labels involved in each path and check for common occurrences. A first measure can be taken by computing the number of labels present in both paths out of the total number of labels. We also add a bonus to this resulting value if the order of labels is the same in both paths. Both values are weighted with the largest part of the weight dedicated to the number of common labels. More formally, this is computed as follows: $((2 * common)/total) * w_1 + (ordered/common) * w_2$, where *common* is the number of common labels, *total* is the total number of labels, *ordered* is the number of ordered labels, and w_1, w_2 are weights. Note that if there are several shortest paths between two states, we compute the path similarity measure for all combinations of two paths (one path from each LTS) and we keep the highest value. The intuition is that we focus on the most similar paths, the other ones could be discarded or not used.

Example. Assume the two paths given in Figure 2. There are three common labels (b, e, c) out of 11 labels, resulting in a first similarity value of $(2 * 3)/11 = 0.55$. We also compute a bonus in that case because among these three common labels, two appear in the same order (b, c), resulting in a value of $2/3 = 0.66$. If we use both values given more importance to the first value (e.g., $w_1 = 0.9$ and $w_2 = 0.1$), we obtain a final path similarity value of $0.55 * 0.9 + 0.66 * 0.1 = 0.56$. Our experience showed that the bonus should be reasonable, otherwise it takes too much importance on the final result and makes it irrelevant.

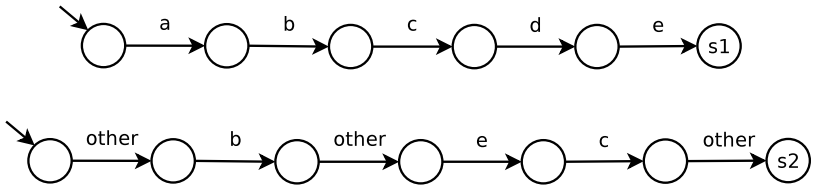

Figure 2: Example for the Path Similarity Criterion.

3.3. Local Criteria

These criteria aim at comparing two states $s_1 \in LTS_1$ and $s_2 \in LTS_2$ by looking at their transitions and states (nature and neighbours). We consider four local criteria:

- counting the number of matching outgoing transitions
- counting the number of matching incoming transitions
- comparison of nature of states (initial or not)
- comparison of neighbour states

Given two sets of transitions T_1 and T_2 outgoing from states s_1 and s_2, respectively, we compute the similarity of those transitions (m_{out}) as follows: ((number of matching transitions in T_1) / $|T_1|$ + (number of matching transitions in T_2) / $|T_2|$) / 2. This measure is undefined if there is no outgoing transitions. The same measure is computed for incoming transitions.

Example. We illustrate with two simple examples where we compare the transitions outgoing from two states $s_1 \in LTS_1$ and $s_2 \in LTS_2$ (Figure 3). Consider first the two states on the left hand side of Figure 3. We obtain

$m_{out} = ((1/1) + (1/2))/2 = 0.75$ because the transition outgoing from s_1 has a counterpart whereas only one of the two transitions outgoing from s_2 has a matching transition. If we now look at the second example on the right hand side of Figure 3, we have $m_{out} = ((1/2) + (1/2))/2 = 0.5$ because from s_1 (s_2, respectively), only half of the transitions have a match.

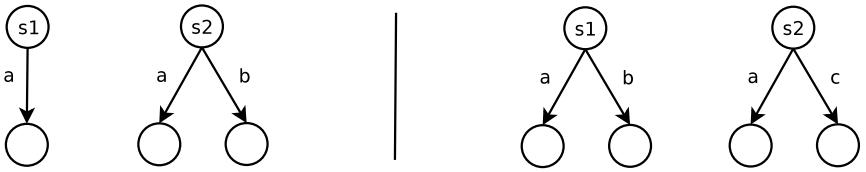

Figure 3: Example for the Transition Matching Criterion.

The nature of two states is simple. If both states are initial or not, we return 1. Otherwise, they have a different nature (one is initial, the other is not), and in that case, we return 0.

The fourth metric takes into account the similarity of neighbour states. The neighbours of a state are all its successor and predecessor states. More precisely, given two states s_1 and s_2, this similarity measure (m_{neig}) is obtained by comparing the predecessors of both states, by comparing the successors of both states, and by computing the average of all these similarity measures.

Example. Suppose two states $s_1 \in LTS_1$ and $s_2 \in LTS_2$ as depicted in Figure 4. The similarity of neighbour states is computed as follows: $m_{neig}[s1, s2] = (m[s1', s2'] + m[s1'', s2''] + m[s1'', s2'''])/3$.

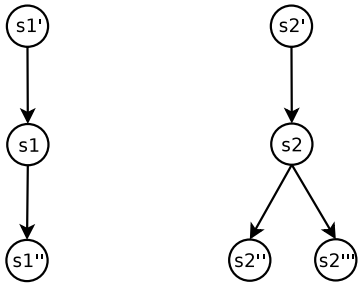

Figure 4: Example for the Neighbour Similarity Criterion.

Since the computation of state similarity uses the matrix itself (for neighbour states), we use an iterative algorithm that stops when the matrix sta-

bilizes. In practice, the iterative process terminates when the distance δ between two versions of the matrix goes below a fixed threshold. The distance between two matrices is obtained by computing the arithmetic mean of the difference of the two same states in each matrix.

Note that the computation of the matrix always converges to a unique similarity matrix. This convergence can be proven as achieved in [18] by using Banach's fixed point theorem.

3.4. Global Measure

Once the similarity matrix is computed, there are several ways for computing a global score out of the matrix. We present in this section four options:

- Average of all scores in the matrix (named "all"). This score is usually rather low since it is somehow *brute force*, and keeps in the calculation all values, even irrelevant ones between completely different states.

- Average of the best score for each row and for each column (named "best"). This score adopts an optimistic point of view and makes an attempt at matching states with high scores.

- Average of scores above a threshold (named "threshold"). In that case, we keep all values in the matrix above a threshold (e.g., 0.8), assuming that these states are the best matches yet there might be several possible matches.

- Average of a percentage of the best scores in the matrix (named "percentage"). As an example, we can compute the average of the 20% of best scores. This option aims at measuring whether the best values are high or not. There might be a matrix with a good score for average above threshold, but rather low score if we focus on the 20% of best scores because most values in the matrix are low. Suppose for example that there are the following 10 numbers in the matrix: 0.1, 0.2, 0.2, 0.3, 0.35, 0.4, 0.4, 0.45, 0.5, 0.9. Threshold would return 0.9 because there is a single value above 0.8. However, percentage would return 0.7 because we take 20% of the best values (0.5 and 0.9 concretely) and we compute the average.

Example. Table 1 gives an example of matrix obtained applying the aforementioned computations. The "all" global measure indicates that non-bisimilar states are similar to 38%. This score is rather low due to the numerous low values appearing in the matrix. The next global values get rid somehow of states s4, s5 and s6, which clearly have no counterpart in the other LTS. The "best" measure returns a similarity percentage of 80% for non-bisimilar states. The "threshold" measure returns a similarity percentage of 95% (with 0.8 as threshold). This value is very high because there are only 3 values above 0.8 and these values are actually higher than 0.9. The "percentage" measure returns a global mesure of 70% (considering 30% of the best scores). This value is rather low compared to "best" and "threshold" because it takes into account more (low) values. If we use the 20% and 10% of best scores, respectively, we obtain 78% and 95% as global similarity measure, respectively.

	s0	s1	s2	s3	s4	s5	s6
s0	0.9	0.05	0.07	0.08	0.11	0.1	0.19
s1	0.18	0.99	0.35	0.32	0.3	0.3	0.36
s2	0.17	0.36	0.98	0.43	0.38	0.36	0.43
s3	0.15	0.32	0.42	0.77	0.58	0.41	0.48

Table 1: Example for the Global Similarity Measure.

3.5. Tool Support

The partition refinement algorithm and the similarity measures (matrix and global measure) are computed via a tool (DLTS) we implemented in Python, see Figure 5 for an overview. It takes as input two LTSs specified in the textual 'aut' format. In practice, we use the LNT process algebra [4] for specifying high-level concurrent systems and compile these specifications to LTSs in 'aut' format by using CADP compilers [6].

We applied our tool to about 50 examples. Each example consists of a series of LTS files (at least two and sometimes more than ten) in 'aut' format. Each file represents a variant of the first LTS. As a result, there are about 200 'aut' files in our repository of examples. During our experiments, we compared all the files in a given series. It does not make sense to compare two files in two different series because the results would be so different that it would not be helpful for validating our approach. Among these 50 examples,

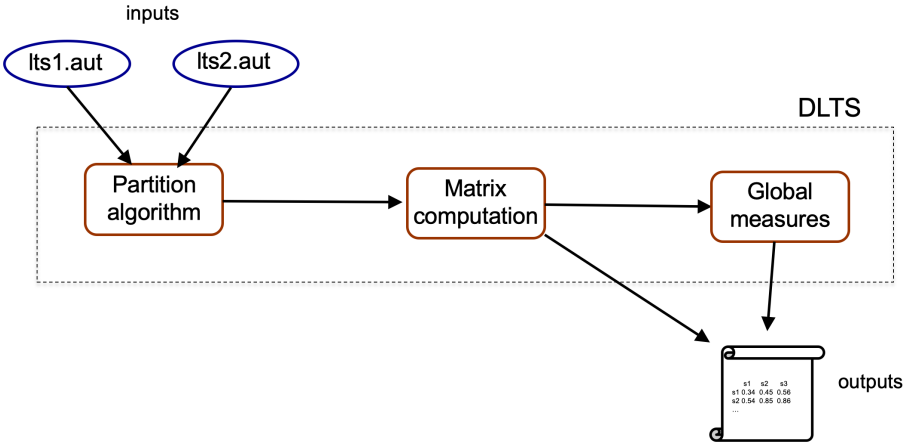

Figure 5: Tool Support.

most of them have been handwritten and some of them have been taken from existing application areas (Internet of Things, Web services, mutual exclusion protocols).

Experiments were carried out to evaluate the quality of the results using the well-known precision and recall measures. In this specific context, we consider the matching of states as basis for these measures. According to our approach, two states are a correct match if a state cannot obtain a higher result in the similarity matrix with any other state. For instance, in Table 1, the couple (s_0,s_0) is a correct match because the state s_0 (row) obtains its higher value (0.9) with s_0 (column). Precision computes the number of correct matched states out of all matched states detected by our approach. This allows us to verify that we do not have too many false positives (irrelevant matched states). Recall corresponds to the number of correct matched states detected by our approach out of all expected matched states. This allows us to measure the number of matched states our approach fails to identify. We have computed the precision and recall measures for several examples taken from our repository. This was not possible to compute these measures for all examples of our database because this requires human expertise and the number of combinations (given 200 examples) is too high. We observed that precision and recall measures are very high (close to 100%) for LTSs which are pretty similar, showing in that case the good quality of our approach. However, when the two compared LTSs are rather different, these values tend to decrease and even to be low for very different LTSs.

As far as performance is concerned, it takes less than a second to compute the matrix (and all global measures) for LTSs involving tens of states and transitions. However, it takes several minutes to compute these outputs for LTSs involving thousands of states and transitions. This comes from both the computation of the partition refinement algorithm and from the matrix computation. The good point is that our approach does not target large LTSs but rather small ones as we will show in the next section with two real-world case studies taken from different application domains.

The tool and all the examples used for validation purposes are available online [27].

4. Applications

In this section, we illustrate with two possible applications of the measure of similarity between LTSs, namely the design of IoT applications and the matching of business process models.

4.1. Composition of IoT Objects

When composing devices and software (object for short in the rest of this section) for building IoT application, each object must exhibit the actions it can execute as well as the order in which these actions must be triggered. Such a public interface can be described using an LTS, as proposed in [10, 9], where labels on transitions correspond to these actions. Two objects interact one with another by synchronizing on same action name (synchronous binary communication model).

Given such a behavioural model for objects, the overall objective is to build a satisfactory composition of objects that satisfies a given goal. The goal is an abstract specification of what the user expects from the resulting composition. It can be modelled using an LTS too using interactions as labels (synchronization of actions) as suggested in [5]. A composition is satisfactory if it satisfies the goal. This can be verified using first the synchronous product to build a unique LTS out of a set of object LTSs, and then comparing the resulting LTS with the goal LTS using equivalence techniques (strong bisimulation here).

This case study aims at building a new IoT application for home security and more precisely for home intrusion detection. The goal of this application is given in Figure 6 using an LTS, which indicates that when a move is

detected, the camera is turned on, an alert message is sent to a mobile phone, and the light in the house is switched on.

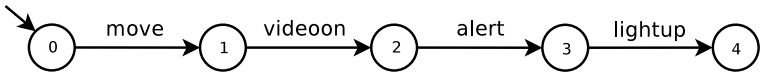

Figure 6: Case Study: Goal.

Usually, in order to build an application satisfying these requirements, the end-user needs a recommender system listing all objets available nearby with their interface. We assume here that after this task the four objects given in Figure 7 are selected as possible candidates. There is first a security sensor that detects movement in the house. When a movement is detected it turns the video on and sends an alert message to a mobile phone. Once the alert is over, the sensor reinitializes and turns the video off. The second object is a connected light that can be repeatedly switched on and off. The third object is a security camera whose video can be activated or not. When activated, watching the video is possible. Finally, the final object is a home security app that can be installed on a smartphone. This app is triggered when receiving an alert. Then, there are several functionalities available for the user such as watching the video or switching the light on/off. Once the alert is over, the app allows the user to initialize it again.

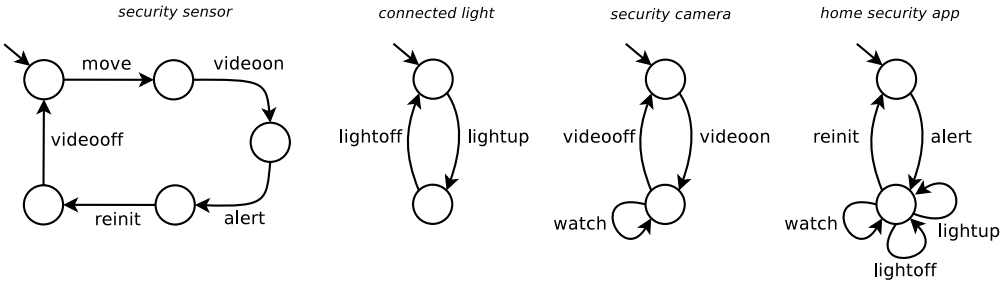

Figure 7: Case Study: Objects.

The next question is the following: are we sure that this selection of objects does satisfy the given abstract goal? This is when equivalence checking comes into play. From the objects given in Figure 7, we can build the resulting LTS using the classic parallel composition operator available in process algebra or the synchronous product of communicating automata. Here, we synchronize two objects on same actions. If an action in one object does not

have any counterpart in another object, this is an independent evolution. As a result we obtain the LTS (generated with CADP) depicted in Figure 8.

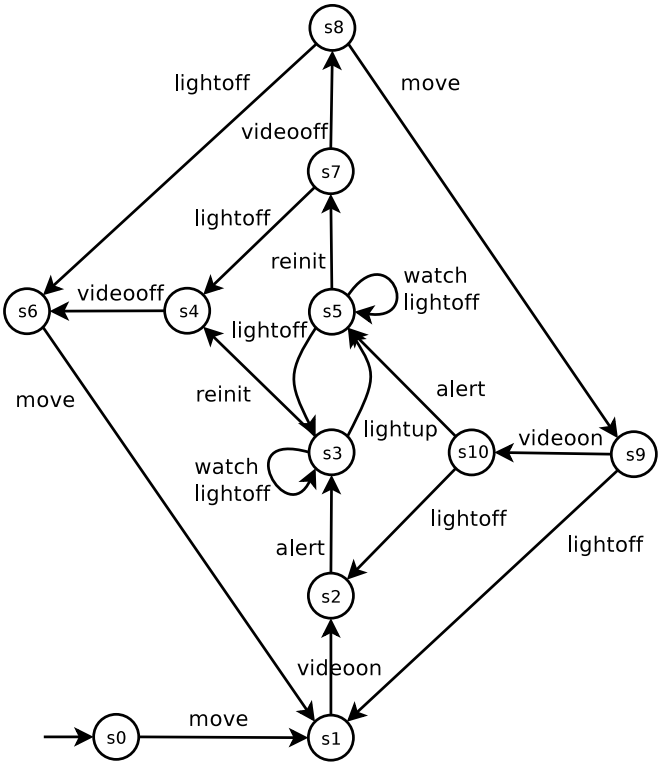

Figure 8: Case Study: Composition LTS.

We now compare both LTSs (goal, Figure 6, and composition LTS, Figure 8) using the DLTS tool. Table 2 shows the resulting matrix obtained after four iterations. The comparison method also indicates that all states are non-bisimilar. The global similarity measure (average of best scores) returns a value of 76%, indicating that both LTSs are not totally different and exhibit portions of their behaviours that are very similar. By looking more carefully at the matrix given in Table 2, we can see that the first three states are very similar with values higher than 80%. But then, most values are very low. If we look at the states in the composition LTS, we understand that this is due to actions present in the objects and their composition that are not taken into account by the abstract goal (e.g., lightoff, videooff).

There are now several options. One can try another selection and com-

	s0	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10
s0	0.97	0.18	0.17	0.17	0.16	0.17	0.47	0.16	0.35	0.22	0.17
s1	0.18	0.86	0.36	0.33	0.3	0.31	0.3	0.3	0.29	0.68	0.33
s2	0.17	0.38	0.88	0.43	0.37	0.39	0.35	0.36	0.34	0.33	0.72
s3	0.19	0.34	0.45	0.77	0.44	0.59	0.4	0.43	0.4	0.37	0.38
s4	0.4	0.42	0.48	0.58	0.62	0.74	0.55	0.58	0.52	0.5	0.48

Table 2: Case Study: First Similarity Matrix.

bination of objects. Another option is to go further in the analysis and comprehension of the current solution. As far as the latter is concerned, we can decide to refine the goal by integrating the missing actions, or we can keep the abstract goal as is and hide in the composition LTS these irrelevant actions. We decide to go for this final option by hidding the actions where light and camera are switched off (lightoff and videooff). Figure 9 shows the resulting LTS (generated with CADP) where hidden actions have been removed for the sake of readability.

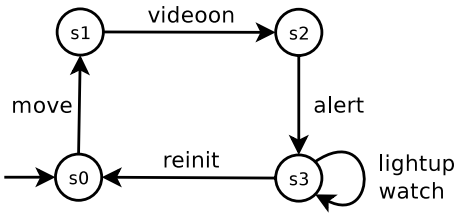

Figure 9: Case Study: Composition LTS (V2).

Table 3 gives the similarity matrix computed by comparing the goal (Figure 6) with this second composition LTS (Figure 9). All states are non-bisimilar but the global measure (average of best scores) increases to 89%. The matrix shows very similar states for the three first states (s0, s1, s2) and lower values for the remaining states (s3, s4). This is quite normal because the two LTSs exhibit several differences in states s3 and s4: (i) the goal is non-looping whereas all objects can loop forever and so the composition LTS, (ii) the reinit and watch actions were not made explicit in the goal but they totally make sense, and (iii) the lightup action is in sequence in the goal whereas it can be repeated in the composition. It is worth noting that these differences make the equivalence not satisfied, but once better understood using our measures, there are no real problems from a functional perspective. Therefore, although the two LTSs are not bisimilar, the end-user could

be satisfied by the proposed composition of objects and accept it as a correct solution for his/her application.

	s0	s1	s2	s3
s0	0.9	0.18	0.17	0.15
s1	0.05	0.99	0.36	0.32
s2	0.07	0.35	0.98	0.43
s3	0.08	0.33	0.44	0.79
s4	0.26	0.37	0.47	0.72

Table 3: Case Study: Second Similarity Matrix.

Last but not least, the user can take advantage of the first LTS generated for the initial composition (Figure 8) where all possible executions are enumerated to verify additional properties (absence of deadlocks, a certain action is always reachable, some action occurs after another one, etc.). This can be achieved using temporal logic and model checking (we use MCL [14] and the CADP model checker [6], respectively, in this work).

4.2. Matching of Process Models

Business process modeling (BPM) is the activity of representing processes of an enterprise, so that the current process may be understood, analysed and improved. Business Process Model and Notation (BPMN) is a graphical representation for specifying business processes. The latest version is BPMN 2.0 and was published as an ISO standard a few years ago [7]. BPMN is a workflow-based graphical notation and is now being used by most business experts and analysts.

In this context, one of the central issues aims at providing techniques for comparing two versions of a process model. Comparing processes is useful for tackling several problems such as process reconfiguration or evolution [12], process harmonization [31] or effective search [13]. Comparison of process models involves matching, that is, the detection of correspondences between activities. In this section, we will show how the computation of bisimilar states help to detect such correspondences. Our measure of similarity goes farther by also focusing on the differences between both models and quantifying these differences.

We use as example in this section a process modelling an online booking system (Figure 10), similar to an example presented in [22]. Note that in

BPMN, X within a diamond states for an exclusive choice. In the figure, the first occurrence of this choice corresponds to a join pattern whereas the second occurrence corresponds to a split pattern. In this BPMN process, a client can first submit a request and receive an answer. Depending on this answer, (s)he can either decide to submit another request, choose to quit (abort), or make a booking. In this last case, the client also makes a payment to confirm his/her booking. This BPMN process can be automatically transformed to an equivalent LTS using existing tools such as VBPMN [11, 12].

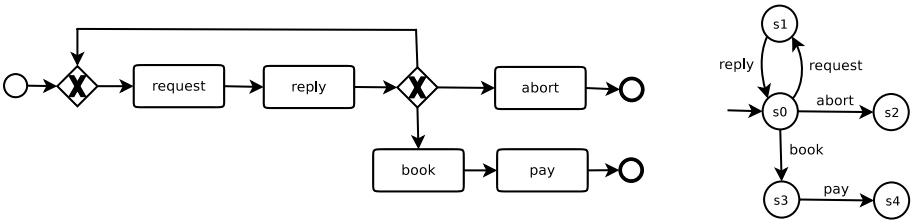

Figure 10: Process Model (V0): BPMN Process (left) and LTS Model (right).

Assume we want to improve this first process by using a connected mode in order to identify precisely the client and thus simplify the payment task. As a result, we obtain the process given in Figure 11 where the first activity corresponds now to this identification step (login).

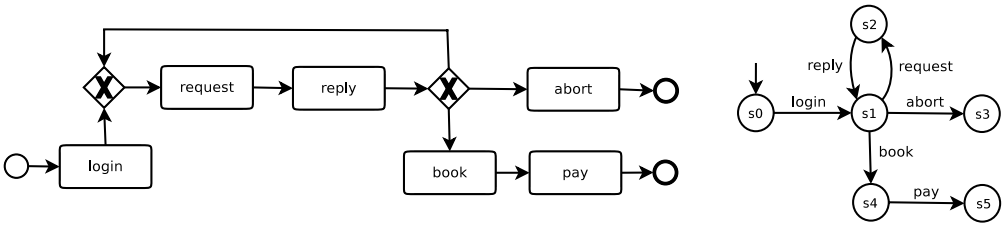

Figure 11: Process Model (V1): BPMN Process (left) and LTS Model (right).

When we compare these two versions of the BPMN process, our similarity measure indicates that just one LTS (the one for V1) has non-bisimilar states. More precisely all states in the LTS for the V0 process have bisimilar counterparts in the LTS for the V1 process whereas state s0 in V1 LTS is not bisimilar to any state in V0 LTS. This is normal because this initial state with one outgoing transition labelled with "login" is the only difference between these two models.

Let us introduce a new extension of the BPMN process introduced beforehand. Assume that now after deciding to book some good, the application also stores the transaction in a database (log activity) as shown in Figure 12.

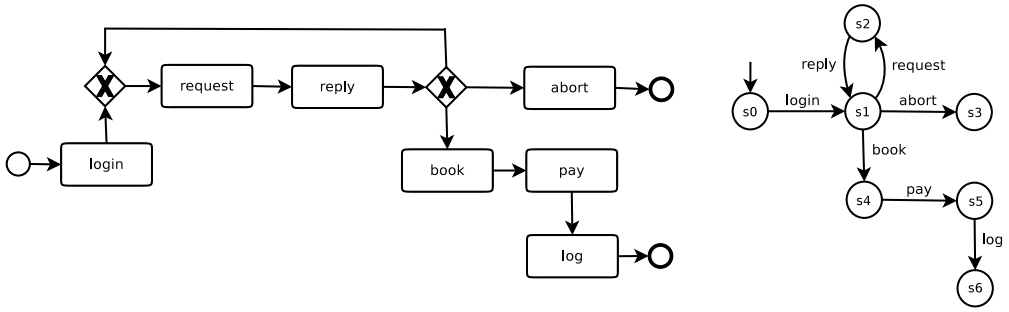

Figure 12: Process Model (V2): BPMN Process (left) and LTS Model (right).

We apply again our tool to compare versions V1 and V2 of the BPMN process. We obtain as result the matrix given in Table 4. As far as global measures are concerned, 69% of states are non-bisimilar and those states are similar to 89% (average of best scores) and to 95% (average of best scores above threshold of 0.8). The matrix shows that most non-bisimilar states (s0, s1, s2, s4) are very similar with high values for couples (si,si) where $i \in \{0, 1, 2, 4\}$. The matrix also highlights the main difference between these two models, consisting of the addition of the "log" activity. Indeed, state s5 in V2 LTS has no counterpart with V1 LTS, and this is why this state only exhibits low values (less than 0.44) in the matrix for all states from V1 LTS.

	s0	s1	s2	s4	s5
s0	0.99	0.3	0.43	0.28	0.29
s1	0.31	0.95	0.45	0.35	0.4
s2	0.43	0.44	0.99	0.53	0.37
s4	0.29	0.42	0.5	0.88	0.44

Table 4: Case Study: Similarity Matrix for BPMN Processes (V1 vs. V2).

Finally, as noticed at the end of Section 4.1, one can make use of model checking tools in order to verify properties of interest on any of the LTS obtained during the comparison process.

5. Related Work

Comparing automata-based models using equivalence techniques is not a new problem. It was studied for instance in the context of the composition of web services, see, e.g. [28, 2]. Observational equivalence was used in [28] for checking that two versions of a service composition were the same. In [2], the authors proposed one compatibility definition based on bisimulation techniques for checking whether two web services can interact properly. In this work, our focus is on quantitative aspects of non-equivalent behavioural models.

In [29], the authors measure the similarity of Labelled Transition Systems (LTSs) with respect to a simulation and a bisimulation notion inspired from the equivalence relations. The measuring techniques use weighted quantitative functions which consist in a simple (not iterative), forward and parallel traversal of two LTSs. This work does not return any global similarity measure and the differences which distinguish one entity from another.

In [15, 17], the authors rely on a similarity flooding algorithm for computing the matrix of correspondences between models. A forward and backward similarity propagation is used in [15] to compare data structures described with directed labelled graphs. However, the tool does not enable a fully automated matching because the user should manually adjust some matches. The match operator introduced in [17] measures the similarity between different versions of software units described using Statecharts. The similarity measuring combines a set of static and behavioural matchings. The behavioural matching is computed using a flooding algorithm and relies on the bisimulation notion presented in [29]. Flooding algorithms were also used for measuring the compatibility of behavioural models of web services in [19, 20]. Our iteration process is very similar to similarity flooding algorithm but tackles the problem with a different angle by focusing only on non-bisimilar states.

The simulation preorder is extended in [3] to a quantitative setting. It presents three notions of distances (correctness, coverage, robustness), which resides in making each player of a simulation game pay a certain price for his/her choices. These distances are comparable to the global measures proposed in this paper. There is no local criterion used in their work.

In [32], the authors present an approach (SpecDiff) to compute the differences between two LTSs obtained by compilation from CSP, representing the evolving behaviors of a concurrent program. SpecDiff considers LTSs as Typed Attributed Graphs (TAGs), in which states and transitions are en-

coded in finite dimensional vector spaces. It then computes a maximum common subgraph of two TAGs, which represents an optimal matching of states and transitions between two LTSs. This approach aims at pairing states and transitions for debugging purposes whereas we analyze the structure of both LTSs without mandatorily finding a match. Moreover, our approach is more general-purpose and not only designed for program debugging.

The approach in [30] aims at comparing state machines in terms of their language (the externally observable sequences of events that are permitted or not), and in terms of their structure (the actual states and transitions that govern the behaviour). The language comparison exploits model-based testing approach. The structure comparison uses what we call local criteria in our paper, by looking at the similarity of surrounding transitions and source/target states. They do not rely on any notion of distances as we did to compare the situation of both states in their respective LTSs. They do not focus on non-bisimilar states only as we do. As far as application is concerned, they apply their approach to reverse-engineering state machines from program traces.

A distance between processes modelled as trees is defined in [23, 24] by computing the costs to transform one of the trees into the other. This notion of distance between processes is defined using coinduction. This approach applies in the case of both finite and infinite trees. The notion of k-bisimulation was introduced in [25]. It considers weak bisimulation and more specifically the weak equivalence notion introduced by Milner in [16]. K-bisimulation measures the number of actions to be hidden for establishing weak equivalence between two processes modelled using LTSs. Thoses measures are less precise than ours since they do not give any detailed measure of distance among the states of both LTSs. It is closer to our global measures of similarity, which gives a rough estimation of how far the two LTSs are one from another.

6. Concluding Remarks

We have presented in this paper a set of metrics that allows us to quantify the difference between the non-bisimilar parts of two LTSs. This similarity measure combines global and local criteria for computing a matrix that compares all non-bisimilar states in both LTSs. The computed matrix is used in a second step for computing global measures of similarity that are helpful to distinguish totally different LTSs and almost bisimilar ones. Our approach

is implemented in a tool and was applied on a set of about 200 LTSs for validating the ideas. Beyond that, we applied our solution to two concrete application areas, namely the design of IoT applications by composition of objects and the matching of business process models. These two case studies show how our similarity measure can be used in practice to solve concrete problems.

As far as future work is concerned, we plan to extend our work to support other notions of bisimulations. Another perspective aims at taking advantage of all the values gathered in the similarity matrix to refine our comprehension of the differences between the two LTSs and extract more outputs from the matrix. Finally, as a long term perspective, we would like to work on the optimization of the tool support to make our approach scalable on large LTSs consisting of possibly millions of states/transitions.

Acknowledgments

We are grateful to Frédéric Lang for his help and expertise on bisimulation and equivalence checking.

References

[1] Baier, C., Katoen, J., 2008. Principles of Model Checking. MIT Press.

[2] Bordeaux, L., Salaün, G., Berardi, D., Mecella, M., 2004. When are Two Web Services Compatible?, in: Proc. of TES'04, Springer. pp. 15–28.

[3] Cerný, P., Henzinger, T.A., Radhakrishna, A., 2010. Simulation Distances, in: Proc. of CONCUR'10, Springer. pp. 253–268.

[4] Champelovier, D., Clerc, X., Garavel, H., Guerte, Y., Lang, F., McKinty, C., Powazny, V., Serwe, W., Smeding, G., 2018. Reference Manual of the LNT to LOTOS Translator (Version 6.7). INRIA/VASY and INRIA/CONVECS, 153 pages.

[5] Durán, F., Salaün, G., Krishna, A., 2019. Automated Composition, Analysis and Deployment of IoT Applications, in: Proc. of TOOLS 2019, Springer. pp. 252–268.

[6] Garavel, H., Lang, F., Mateescu, R., Serwe, W., 2013. CADP 2011: A Toolbox for the Construction and Analysis of Distributed Processes. STTT 15, 89–107.

[7] ISO/IEC, 2013. International Standard 19510, Information technology – Business Process Model and Notation.

[8] Kanellakis, P.C., Smolka, S.A., 1990. CCS Expressions, Finite State Processes, and Three Problems of Equivalence. Inf. Comput. 86, 43–68.

[9] Krishna, A., Pallec, M.L., Mateescu, R., Noirie, L., Salaün, G., 2019a. IoT Composer: Composition and Deployment of IoT Applications, in: Proc. of ICSE'19, Montreal, IEEE / ACM. pp. 19–22.

[10] Krishna, A., Pallec, M.L., Mateescu, R., Noirie, L., Salaün, G., 2019b. Rigorous Design and Deployment of IoT Applications, in: Proc. of FormaliSE'19, ACM.

[11] Krishna, A., Poizat, P., Salaün, G., 2017. VBPMN: Automated Verification of BPMN Processes, in: Proc. of IFM'17, Springer. pp. 323–331.

[12] Krishna, A., Poizat, P., Salaün, G., 2019c. Checking business process evolution. Sci. Comput. Program. 170, 1–26.

[13] Kunze, M., Weidlich, M., Weske, M., 2011. Behavioral similarity - A proper metric, in: Proc. of BPM'11, Springer. pp. 166–181.

[14] Mateescu, R., Thivolle, D., 2008. A Model Checking Language for Concurrent Value-Passing Systems, in: Proc. of FM'08, Springer.

[15] Melnik, S., Garcia-Molina, H., Rahm, E., 2002. Similarity Flooding: A Versatile Graph Matching Algorithm and Its Application to Schema Matching, in: Proc. of ICDE'02, IEEE Computer Society. pp. 117–128.

[16] Milner, R., 1989. Communication and Concurrency. Prentice Hall.

[17] Nejati, S., Sabetzadeh, M., Chechik, M., Easterbrook, S.M., Zave, P., 2007. Matching and Merging of Statecharts Specifications, in: Proc. of ICSE'07, IEEE Computer Society. pp. 54–64.

[18] Ouederni, M., Fahrenberg, U., Legay, A., Salaün, G., 2017. Compatibility Flooding: Measuring Interaction of Services Interfaces, in: Proc. of SAC'17, ACM. pp. 1334–1340.

[19] Ouederni, M., Salaün, G., Pimentel, E., 2010. Quantifying Service Compatibility: A Step beyond the Boolean Approaches, in: Proc. of ICSOC'10, Springer. pp. 619–626.

[20] Ouederni, M., Salaün, G., Pimentel, E., 2011. Measuring the Compatibility of Service Interaction Protocols, in: Proc. of SAC'11, ACM. pp. 1560–1567.

[21] Paige, R., Tarjan, R.E., 1987. Three Partition Refinement Algorithms. SIAM J. Comput. 16, 973–989.

[22] Poizat, P., Salaün, G., 2012. Checking the realizability of BPMN 2.0 choreographies, in: Proc. of SAC'12, ACM. pp. 1927–1934.

[23] Romero-Hernández, D., de Frutos-Escrig, D., 2012. Defining Distances for All Process Semantics, in: Proc. of FMOODS/FORTE'12, Springer. pp. 169–185.

[24] Romero-Hernández, D., de Frutos-Escrig, D., 2014. Coinductive Definition of Distances between Processes: Beyond Bisimulation Distances, in: Proc. of FORTE'14, Springer. pp. 249–265.

[25] Ruvo, G.D., Lettieri, G., Martino, D., Santone, A., Vaglini, G., 2015. k-Bisimulation: A Bisimulation for Measuring the Dissimilarity Between Processes, in: Proc. of FACS'15, Springer. pp. 181–198.

[26] Salaün, G., 2019. Quantifying the similarity of non-bisimilar labelled transition systems, in: Proc. of FOCLASA'19, Springer. pp. 211–225.

[27] Salaün, G., 2020. DLTS Tool and Examples. `http://convecs.inria.fr/people/Gwen.Salaun/Tools/dlts.zip`.

[28] Salaün, G., Bordeaux, L., Schaerf, M., 2004. Describing and Reasoning on Web Services using Process Algebra, in: Proc. of ICWS'04, IEEE. pp. 43–50.

[29] Sokolsky, O., Kannan, S., Lee, I., 2006. Simulation-Based Graph Similarity, in: Proc. of TACAS'06, Springer. pp. 426–440.

[30] Walkinshaw, N., Bogdanov, K., 2013. Automated Comparison of State-Based Software Models in Terms of Their Language and Structure. ACM Trans. Softw. Eng. Methodol. 22, 13:1–13:37.

[31] Weidlich, M., Mendling, J., Weske, M., 2011. A foundational approach for managing process variability, in: Proc. of CAiSE 2011, Springer. pp. 267–282.

[32] Xing, Z., Sun, J., Liu, Y., Dong, J.S., 2011. Differencing Labeled Transition Systems, in: Proc. of ICFEM'11, Springer. pp. 537–552.