Data as processes: introducing measurement data into CARMA models

Data as processes: introducing measurement data into CARMA models

Stephen Gilmore
Laboratory for Foundations of Computer Science
The University of Edinburgh
Edinburgh, Scotland
Stephen.Gilmore@ed.ac.uk

Measurement data provides a precise and detailed description of components within a complex system but it is rarely used directly as a component of a system model. In this paper we introduce a model-based representation of measurement data and use it together with modeller-defined components expressed in the CARMA modelling language. We assess both liveness and safety properties of these models with embedded data.

1 Introduction

A formal model of a real-world system uses abstraction to distill the most important elements of the system into a succinct representation which is amenable to formal reasoning and analysis. If the modeller creating the model has chosen the right level of abstraction for their analysis then the insights which are gained from model-based reasoning are also applicable to the real-world system itself. If, however, some of the important elements of the system have been mis-represented in the model then the insights gained by model-based reasoning and analysis are of no value, no matter how much trouble or care was taken to obtain them from the (flawed) formal model.

For dynamic models used to study performance properties such as throughput, utilisation, and satisfaction of service-level agreements, one challenge which the modeller must face is representing the timed behaviour of systems accurately. Depending on the kind of model that is being created, either continuous sure variables or random variables from a particular random number distribution are used to abstract aspects of timed behaviour in the system under study. These variables are parameters of the model, allowing it to be used in a suite of experiments which explore the behaviour of the model when some of the parameters are perturbed. For such a modelling study to be informative about the system under study it is then necessary to ensure that these parameters are correctly chosen to reflect the durations of the corresponding events in the system.

Techniques for abstracting empirical univariate distributions into statistical distributions such as phase-type distributions are well known and available as algorithms [6] and even as software tools [12]. However, in the case of systems where spatial aspects play a significant role in addition to timed behaviour we have several correlated variables and a multivariate distribution which means that finding a suitable abstraction is not so easy. Where a classical model of a spatially-distributed system would typically use a co-ordinate system to provide an abstract representation of space, our concrete component instead uses literal latitude and longitude co-ordinates to represent the current position of a mobile component. The result is a model which is a mix of abstract components crafted by the modeller and concrete components which have been automatically generated from measurement data. This allows us to build models of systems where we selectively choose not to abstract one component, but instead to represent it literally in order to ensure that we do not misrepresent it via an inappropriate abstraction.

M.H. ter Beek and M. Loreti (Eds.): Workshop on
FORmal methods for the quantitative Evaluation
of Collective Adaptive SysTems (FORECAST'16).
EPTCS 217, 2016, pp. 31–42, doi:10.4204/EPTCS.217.5

From the viewpoint of model-based testing we should see each concrete component as a *black box* component within the model. The component offers up the values of its attributes at any time, but the logic as to why the attribute values change as they do is not represented anywhere in the model, neither in the concrete components generated from measurement data nor in the abstract components defined by the modeller. Measurement data can be easily obtained from an instrumented system and one can often be in the situation of having an embarrassingly large volume of measurement data. Because the concrete components admit no compact representation of their behaviour the modelling formalism which we use must be able to tolerate large unstructured components with real-valued attributes such as latitude and longitude, paired with timestamps.

Many modelling formalisms are not able to meet this challenge. Classical Petri nets, process algebras, and layered queueing networks do not provide the data types and data structures which are needed to represent concrete components within the model. Here however, we are working with CARMA (Collective Adaptive Resource-sharing Markovian Agents) [7] a modern feature-rich modelling language which in addition to providing a stochastic process algebra of guarded recursive processes with unicast and broadcast communication also provides the primitive data types and data structures of a general-purpose programming language. These features are supplemented by encapsulation mechanisms, general function definitions, and iterative constructs for defining collectives of components. Together these features give the modeller sufficient linguistic power to represent concrete components directly within CARMA models, and we utilise this strength of CARMA modelling here.

2 Background

Models in the CARMA language consist of a *collective* of *components*, set in an *environment* representing the context in which the components operate. The collective is a parallel composition of components, each of which consists of a *process* which represents the component's behaviour, and a *store* which represents the component's knowledge.

Stores map *attribute names* to *basic values* of primitive types such as boolean, integer and real. Values such as these can be passed as parameters when processes communicate. An output action $\alpha\langle\vec{v}\rangle$ by one process can be matched with an input action $\alpha(\vec{x})$ by another process provided the length of the vector of values $\vec{v}$ is the same as the length of the vector of variables $\vec{x}$. The arity of a communication action must be consistent throughout the model: it cannot be used to pass one value at one point and two (or more) values at another. Communication actions can either be *unicast* or *broadcast* in CARMA. In all, this provides four types of actions in CARMA:

broadcast output	$\alpha^\star[\pi]\langle\vec{e}\rangle\sigma$	asynchronous (non-blocking) broadcast action α to communication partners identified by the predicate π; send the values of expressions $\vec{e}$ evaluated in the local store γ; then apply the update σ to γ.
broadcast input	$\alpha^\star[\pi](\vec{x})\sigma$	receive a tuple $\vec{x}$ of values $\vec{v}$ sent with an action α from a component whose store satisfies the predicate $\pi[\vec{v}/\vec{x}]$; then apply the update σ to local store γ.
unicast output	$\alpha[\pi]\langle\vec{e}\rangle\sigma$	synchronous (blocking) unicast action α to any communication partner satisfying the predicate π; send the values of expressions $\vec{e}$ evaluated in the local store γ; then apply the update σ to γ.
unicast input	$\alpha[\pi](\vec{x})\sigma$	(point-to-point) receive a tuple $\vec{x}$ of values $\vec{v}$ sent with an action α from a component whose store satisfies the predicate $\pi[\vec{v}/\vec{x}]$; then apply the update σ to local store γ.

The use of predicates to describe communication partners means that CARMA supports the *attribute-based communication* paradigm [2], as found in languages such as SCEL [9], where dynamic collections of components called *ensembles* are formed through having attributes in common. In process algebras such as PEPA [10] where data is abstracted out of the model, attributes are not present and thus attribute-based communication is not possible. Communication partners are determined statically in PEPA whereas they are determined dynamically in CARMA and SCEL.

Processes $(P, Q, \ldots)$ in CARMA are defined by the following grammar:

$$\begin{array}{rcl} P,Q & ::= & \textbf{nil} \mid \textbf{kill} \mid act.P \mid P+Q \mid P|Q \mid [\pi]P \mid A \quad (A \triangleq P) \\ act & ::= & \alpha^\star[\pi]\langle\vec{e}\rangle\sigma \mid \alpha^\star[\pi](\vec{x})\sigma \mid \alpha[\pi]\langle\vec{e}\rangle\sigma \mid \alpha[\pi](\vec{x})\sigma \end{array}$$

By convention in a CARMA model activity names begin with a lowercase letter, function and component names begin with a capital letter, and process names are written in all caps. Expressions in the CARMA language (as used in function bodies) are generated by the following grammar.

$$\begin{array}{rcl} e_1,e_2,e_3 & ::= & \textbf{return}\ e_1 \mid \textbf{if}(e_1)\{e_2\} \mid \textbf{if}(e_1)\{e_2\}\ \textbf{else}\ \{e_3\} \mid e_1;e_2 \mid a_1 \mid b_1 \\ a_1,a_2 & ::= & 0 \mid 1 \mid \cdots \mid -a_1 \mid a_1+a_2 \mid a_1-a_2 \mid a_1*a_2 \mid a_1/a_2 \\ b_1,b_2 & ::= & \textbf{true} \mid \textbf{false} \mid a_1 > a_2 \mid a_1 >= a_2 \mid a_1 == a_2 \mid a_1 <= a_2 \mid a_1 < a_2 \\ & \mid & !b_1 \mid b_1 \,\&\&\, b_2 \mid b_1 || b_2 \end{array}$$

3 Case study: Bus fleet management

For our case study in this paper we show how data on the movement of a bus travelling through the city of Edinburgh can be incorporated into a CARMA model. The purpose of the modelling will be to check whether or not the bus follows the intended route by matching its movements against a high-level description of the route in terms of regions of the city described by predicates. This is an instance of a *fleet management* problem as studied in the transportation modelling community: it is important to know the location of all of the vehicles in the fleet and to know that they are serving their assigned routes. Managing the assignment of buses to routes is not as easy in practice as it might appear: changes of assignment are needed during the working day as problems such as vehicular mechanical failures, road closures, or driver unavailabilty can cause buses to be cancelled or re-routed in ways that would be impossible to predict at the start of the day.

Our specific example is Transport for Edinburgh's Service 100, which travels between Edinburgh airport and Edinburgh city centre. We can characterise this route as having five significant regions: the airport, suburban area 1, suburban area 2, the city centre, and the garage where the bus is parked overnight. These areas are shown in Figure 1, together with a GPS trace of bus fleet number 937 serving this route. The definition of these regions is given in Table 1.

In the dataset that we are working with here, the position of a bus has been registered once every minute. Regions should be chosen to be large enough to make it effectively improbable that a bus can enter the region and exit from it again without having been observed at least once within it. Regions should bound a portion of the bus route, but not so tightly that small measurement errors in GPS readings could cause a bus to be perceived as outside that region. We have chosen our regions to be simple rectangles because it is easy to test whether a point lies within a simple geometric shape such as this. We defined five CARMA predicates to test whether a point lies in a region. These are `AtAirport`, `InSuburbs1`, `InSuburbs2`, `InCentre` and `AtGarage`. The CARMA predicate `AtAirport` is shown in Figure 2. The other predicates are similarly easy to define.

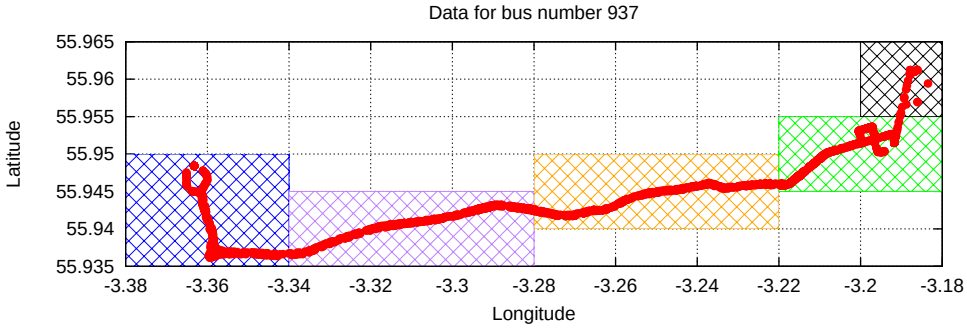

Figure 1: The route of bus number 937 in the fleet, assigned to service 100. The route includes the airport (in the bottom left hand corner, in blue), suburban area 1 (in purple), suburban area 2 (in orange), the city centre (in green), and the garage (in the top right hand corner, in black).

Significant latitudes	*Significant longitudes*	*Region definitions*
$lat_1 = 55.935$	$long_1 = -3.38$	$airport = [(long_1, lat_1), (long_2, lat_4)]$
$lat_2 = 55.940$	$long_2 = -3.34$	$suburbs_1 = [(long_2, lat_1), (long_3, lat_3)]$
$lat_3 = 55.945$	$long_3 = -3.28$	$suburbs_2 = [(long_3, lat_2), (long_4, lat_4)]$
$lat_4 = 55.950$	$long_4 = -3.22$	$centre = [(long_4, lat_3), (long_6, lat_5)]$
$lat_5 = 55.955$	$long_5 = -3.20$	$garage = [(long_5, lat_5), (long_6, lat_6)]$
$lat_6 = 55.965$	$long_6 = -3.18$	

Table 1: Table of region definitions in terms of latitude and longitude coordinates.

```
fun bool AtAirport(real long, real lat) {
  if (long > long1 && long < long2 && lat > lat1 && lat < lat4) {
    return true;
  } else {
    return false;
  }
}
```

Figure 2: The `AtAirport` function in CARMA.

3.1 Generating a concrete component from measurement data

Given measurement data whose records consist of latitude and longitude coordinates together with a timestamp it is straightforward to generate a concrete component which introduces this measurement data into our CARMA model. We generate a straight-line process which broadcasts each *move* action as it occurs. We choose broadcast output because in general we do not want the abstract components of the model to alter the behaviour of the concrete components. The parameters of the *move* action are the measurement data in the form of a five-tuple $\langle latitude, longitude, hour, minutes, seconds \rangle$. This conversion from measurement data into a CARMA component is performed automatically with a Python script. Figure 3 illustrates this process. The updates σ_0, σ_1 and σ_2 are the obvious updates of the local store to hold the current values of latitude, longitude, hours, minutes, and seconds.

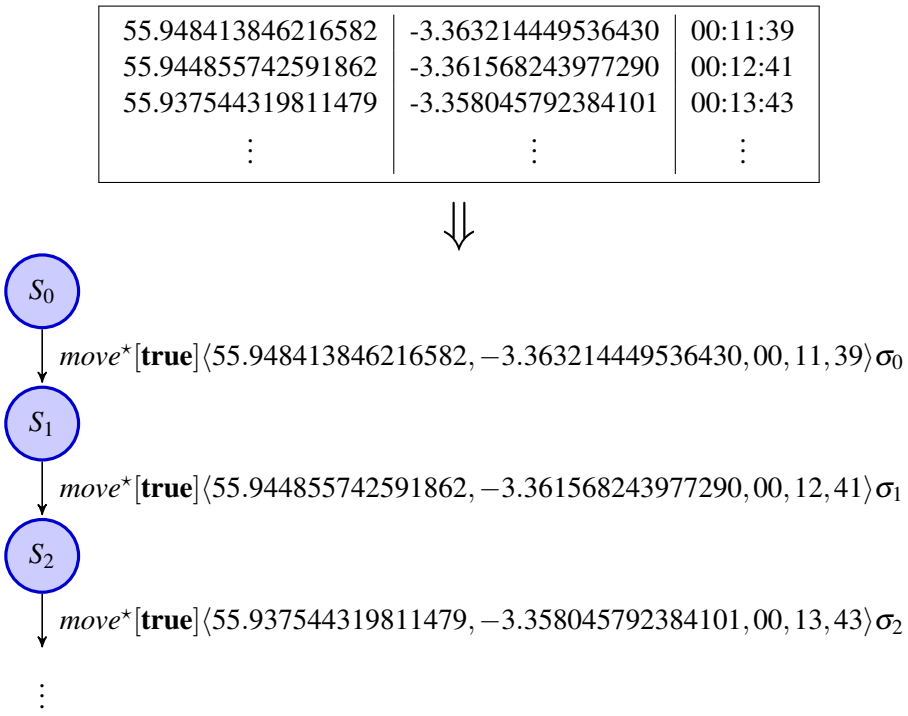

Figure 3: Converting measurement data into concrete components in our CARMA model.

Now that we have our measurement data within our CARMA model we can use the CARMA Eclipse Plugin to execute the model and investigate its behaviour using a *measure* defined in the CARMA model. Measures are real-valued functions which compute some result of the current state of the model, allowing this to be visualised as an assessment of the model's behaviour. We can define a measure in CARMA as shown below.

measure *MaxLatitude* = *max* { **my**.*latitude* };

The *Bus* component which we have generated has an attribute in its local state named *latitude* which is updated after every movement action. A component refers to its own local state using the prefix **my** in CARMA (much like the use of **this** in Java). The particular measure shown above records the maximum value of the latitude seen at all timepoints along the trace of the *Bus* component as it executes. The results are shown in Figure 4.

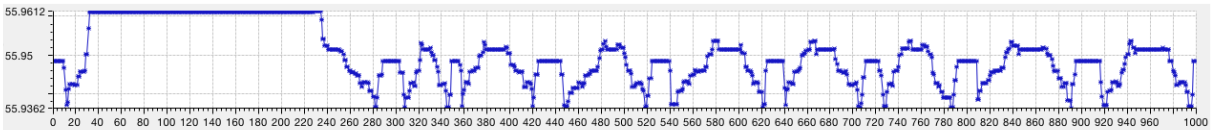

Figure 4: A plot generated by the CARMA Eclipse Plugin showing the maximum value of the *latitude* attribute of the *Bus* component in our model. The long constant high value of latitude near the start of the trace records the bus being parked in the garage overnight.

This plot assures us that the model exhibits *some* behaviour, in that the latitude of the bus is changing, but we do not yet know whether or not it is following the route of the 100 service, or staying within the five regions of interest specified earlier. We will define an additional component to investigate these questions.

3.2 Defining a probe to monitor the concrete component

The next step in checking the correctness of a bus journey is to be able to monitor its behaviour by adding an abstract component (defined by the modeller) to ensure that the progress of the bus moving between regions is as we expect, and that the bus does not leave the five regions which we have defined (see Table 1).

We use the term *probe* for a component whose purpose is simply to monitor changes in another component [5, 8]. The function of a probe is to make it convenient to express checkable properties of a model. A probe is a finite-state automaton which recognises the language of acceptable state transitions within a model and rejects all unacceptable state transition sequences. Probes are sometimes expressed in formal language terms as regular expressions and sometimes as timed automata [4, 3].

Probes can be used to check both safety and liveness properties of models. Here we are interested in two liveness properties and one safety property, as described below.

Liveness 1: The bus visits the airport (probe reaches state `AIRPORT`).

Liveness 2: The bus visits the city centre (probe reaches state `CENTRE`).

Safety: The bus does not leave the five defined regions (probe never reaches state `ERROR`).

Perhaps the most natural description of the journey of the bus would be to separate out the Airport journey (from the airport to the city centre) and the Return journey (from the city centre to the airport). Denoting these A and R respectively, we would note that the journey from the airport to the city centre passes through the two suburban regions in the order $[S_1^A; S_2^A]$ whereas the return journey passes through the two suburban regions in the order $[S_2^R; S_1^R]$. The probe which captures this separation of the Airport and Return journeys is presented in Figure 5. The predicate guards which label each arrow have been omitted to reduce clutter. The selection of start state means that this probe can only be applied to vehicles which begin their journey at the airport. State E indicates that an error has occurred.

Although this description of the probe is perfectly correct from the abstract notion of the bus route it is in practice rather too unforgiving of measurement errors. For a bus stopped on the border between the region $suburbs_1$ and the region $suburbs_2$ a small error in GPS measurement could cause the sequence of observations $[S_1^A; S_2^A; S_1^A; S_2^A]$ to be seen, and this cannot be accepted by the probe presented in Figure 5.

For this reason, we work with a looser specification of the bus route, as described by the probe in Figure 6. This does not differentiate between the Airport and Return routes and has the dual benefits of being more compact and tolerating errors in GPS measurement at the boundaries between regions. For

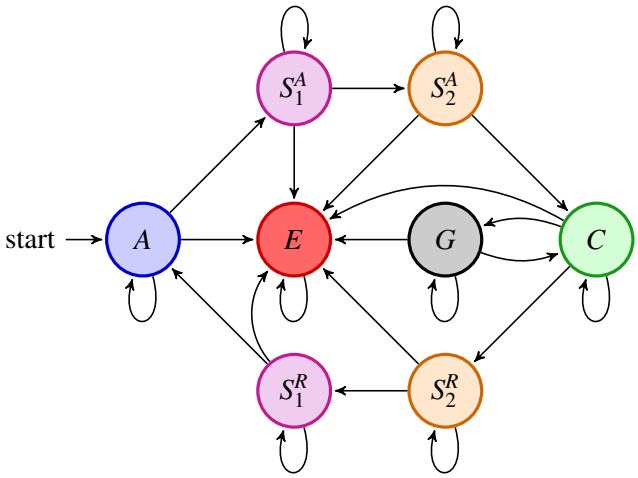

Figure 5: A probe component which separates the outward and return routes.

example, the sequence of observations $[S_1;S_2;S_1;S_2]$ can be accepted by this component. The predicate guards which label each arrow have again been omitted in this diagram to reduce clutter.

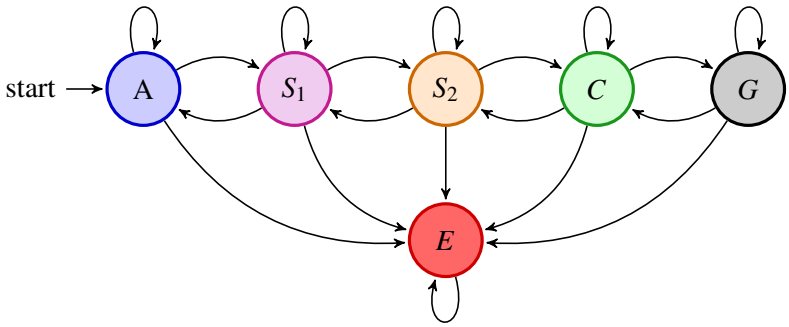

Figure 6: A probe component which does not differentiate between the outward and return routes.

Of course, when this probe is expressed as a CARMA component it is necessary to be absolutely specific about the predicate guards on each transition. The overall effect of the predicates is to track the location of the bus on the basis of its reported latitude and longitude. A transition to the ERROR state of the probe may only be taken if no other outgoing transition from a state is possible. We use the predicates `AtAirport`, `InSuburbs1`, `InSuburbs2`, `InCentre` and `AtGarage` as described previously. The text of the probe as a CARMA component is shown in Figure 7.

3.3 Computing liveness and safety properties using probes

Now we are in a position to be able to use CARMA's measures to interrogate the probe to see states which it visits. In order to turn our probe's observations into numerical measures we count the number of probes in each state. These measures will only ever return 0 or 1 as their results with 1 indicating that

```
component Probe(process Z) {
  store {}

  behaviour {

    AIRPORT = move*[AtAirport(long, lat)](lat,long,h,m,s){}.AIRPORT
            + move*[InSuburbs1(long, lat)](lat,long,h,m,s){}.SUBURBS1
            + move*[!AtAirport(long, lat) &&
                    !InSuburbs1(long, lat)](lat,long,h,m,s){}.ERROR;

    SUBURBS1 = move*[AtAirport(long,lat)](lat,long,h,m,s){}.AIRPORT
            + move*[InSuburbs1(long, lat)](lat,long,h,m,s){}.SUBURBS1
            + move*[InSuburbs2(long, lat)](lat,long,h,m,s){}.SUBURBS2
            + move*[!AtAirport(long, lat) &&
                    !InSuburbs1(long, lat) &&
                    !InSuburbs2(long, lat)](lat,long,h,m,s){}.ERROR;

    SUBURBS2 = move*[InSuburbs1(long,lat)](lat,long,h,m,s){}.SUBURBS1
            + move*[InSuburbs2(long, lat)](lat,long,h,m,s){}.SUBURBS2
            + move*[InCentre(long, lat)](lat,long,h,m,s){}.CENTRE
            + move*[!InSuburbs1(long, lat) &&
                    !InSuburbs2(long, lat) &&
                    !InCentre(long, lat)](lat,long,h,m,s){}.ERROR;

    CENTRE = move*[InSuburbs2(long, lat)](lat,long,h,m,s){}.SUBURBS2
            + move*[InCentre(long, lat)](lat,long,h,m,s){}.CENTRE
            + move*[AtGarage(long,lat)](lat,long,h,m,s){}.GARAGE
            + move*[!InSuburbs2(long, lat) &&
                    !InCentre(long, lat) &&
                    !AtGarage(long, lat)](lat,long,h,m,s){}.ERROR;

    GARAGE = move*[InCentre(long, lat)](lat,long,h,m,s){}.CENTRE
            + move*[AtGarage(long,lat)](lat,long,h,m,s){}.GARAGE
            + move*[!InCentre(long, lat) &&
                    !AtGarage(long, lat)](lat,long,h,m,s){}.ERROR;

    ERROR  = move*[true](lat,long,h,m,s){}.ERROR;
  }

  init { Z }
}
```

Figure 7: The probe from Figure 6 represented as a CARMA component.

the state (AIRPORT, SUBURBS1, ...) has been visited. We add these measures to our model.

```
measure ProbeInStateAIRPORT = #{ Probe[AIRPORT] | true };
measure ProbeInStateSUBURBS1 = #{ Probe[SUBURBS1] | true };
measure ProbeInStateSUBURBS2 = #{ Probe[SUBURBS2] | true };
measure ProbeInStateCENTRE = #{ Probe[CENTRE] | true };
measure ProbeInStateGARAGE = #{ Probe[GARAGE] | true };
```

measure *ProbeInStateERROR* = #{ *Probe*[*ERROR*] | **true** };

There are eleven buses from the Transport for Edinburgh fleet which serve the Airport route at any time. For the day of data which we processed here, these are fleet numbers 937, 938, 939, 940, 941, 943, 945, 947, 948 and 950. We began with bus number 937 and used the CARMA Eclipse Plugin to check our probe against the trajectory of this bus. This showed that the two liveness properties were satisfied (the bus visits the airport and the city centre) and that the safety property was also met (the ERROR state of the probe is never reached). The output from the CARMA Eclipse Plugin is shown in Figure 8.

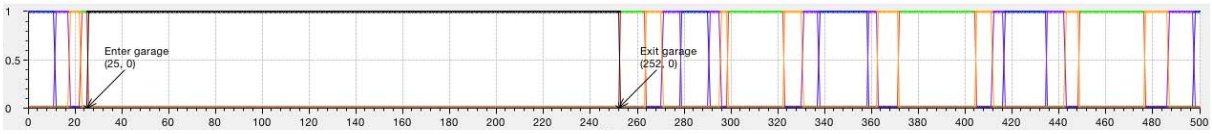

Figure 8: The route of bus number 937 in the fleet. The bus is initially at the airport and enters the garage shortly after midnight. The following day the bus performs the expected route between the city centre and the airport. The ERROR state of the probe is never reached for bus 937.

After this, we applied the same analysis to the remaining ten buses which were serving the 100 route, compiling these results into Table 2. All but one of these buses passed both the liveness tests and the safety test. The bus which failed these tests was bus 947, which fails both of the liveness tests and also fails the safety test.

Fleet number	Initial state	Final state	AIRPORT visited	CENTRE visited	ERROR seen	Probe 100 result
937	AIRPORT	GARAGE	Yes	Yes	No	Accepted
938	GARAGE	AIRPORT	Yes	Yes	No	Accepted
939	GARAGE	CENTRE	Yes	Yes	No	Accepted
940	SUBURBS1	CENTRE	Yes	Yes	No	Accepted
941	CENTRE	GARAGE	Yes	Yes	No	Accepted
943	GARAGE	GARAGE	Yes	Yes	No	Accepted
944	SUBURBS2	GARAGE	Yes	Yes	No	Accepted
945	CENTRE	GARAGE	Yes	Yes	No	Accepted
947	GARAGE	ERROR	No	No	Yes	Rejected
948	GARAGE	AIRPORT	Yes	Yes	No	Accepted
950	GARAGE	SUBURBS1	Yes	Yes	No	Accepted

Table 2: Results of checking our probe against traces from different buses serving the 100 route.

Looking at the results from the CARMA Eclipse Plugin shown in Figure 9, the bus is initially in the garage (probe state is `GARAGE`) but immediately violates the allowable conditions on its latitude and longitude coordinates on leaving the garage (probe state is `ERROR`). The error state of the probe is an absorbing state so once the probe has entered this state it will not escape to any non-error state even if the bus later corrects its position to rejoin the correct route for the service.

Our method of compiling measurement data into concrete components and evaluating it with a probe component has had the desired outcome of finding erroneous behaviour in an unlabelled collection of correct and incorrect trajectories. Bus 947 has been identified as having diverged from the expected route. We now look at its trajectory as a latitude-longitude trace and see if we can conjecture what happened to cause this deviation. Comparing the position with a map of the city of Edinburgh we see that the bus has

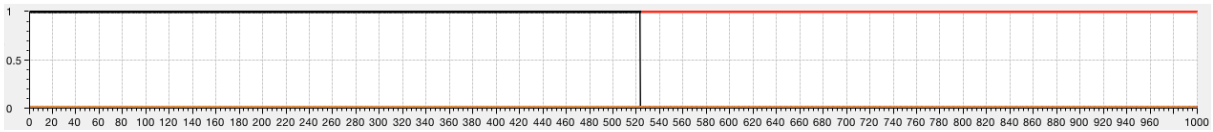

Figure 9: The route of bus number 947 in the fleet. The bus is initially in the garage at the beginning of the trace but enters the ERROR state immediately on exiting the garage.

taken a route away from the city centre towards the coast. A second Transport for Edinburgh garage is located here, and we can conjecture that this bus had a fault or needed some maintenance activity before it was able to serve the 100 route. After visiting the second garage the (now, presumably, repaired) bus returns to the city centre and begins its service from there, as detailed in Figure 10.

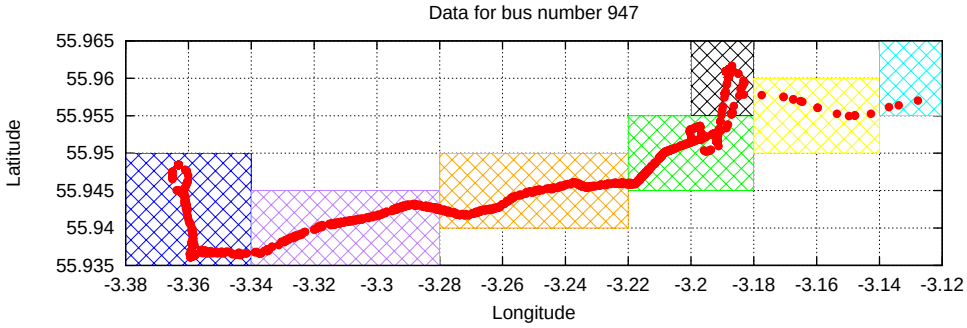

Figure 10: The route of bus number 947 in the fleet, assigned to service 100. The route includes the expected locations of the airport (in the bottom left hand corner, in blue), suburban area 1 (in purple), suburban area 2 (in orange), the city centre (in green), and the garage (in black), as well as the unexpected locations of suburban area 3 (in yellow) and garage 2 (in the top right-hard corner, in cyan).

3.4 Practicality of the method

In our case study here we used GPS measurement data from bus journeys to build our concrete components. The position of the bus is sampled every minute of a twenty-four hour period, thus giving us approximately 1,440 timestamped records of the latitude and longitude of the bus. The measurement data is compiled into a CARMA model by a Python script; this CARMA model is 14,546 lines long. This CARMA model is then compiled into a Java application by the CARMA Eclipse Plugin; this Java application is 81,162 lines long.

The current compilation strategy employed by the CARMA compiler is to compile a CARMA component into a single Java method. For hand-built components created by the modeller this approach has no significant disadvantages but for large generated concrete components this approach runs the risk of overflowing the maximum Java method size of 65,534 bytes. If working with very large measurement data sets, over longer time periods or with finer sampling granularity, then it would be necessary to change the CARMA compilation strategy to compile CARMA processes into individual methods instead and have the compilation images of CARMA components call these methods, thereby moving the problem of maximum method size to reappear again only for CARMA processes which have many attributes and very large update blocks.

4 Related work

In this work we have considered using data concretely as process components in a model. In earlier work, van der Aalst *et al* have devised algorithms for extracting compact process descriptions from data such as system event logs. These algorithms are necessarily incomplete and cannot always find a compact process representation which faithfully encodes an expansive event log. Nonetheless, these *workflow mining* [1] and *process mining* [13] approaches give valuable insights into large data sets by identifying likely causal relations between events and variants of the α algorithm which underlies the workflow mining approach are able to rediscover large classes of processes from event logs.

The *Traviando* simulation trace analyser can also be applied to inverse problems like these, in that it can be used to generate so-called *likely invariants* from a finite execution trace. These likely invariants can then be used to help formulate a compact model of a process which would generate such an event trace [11].

5 Conclusions

We have shown a method of checking liveness and safety properties of CARMA models in which some components of the system which is being modelled are represented without abstraction, using concrete components which are generated automatically from data. The properties which can be checked are those which can be expressed by finite-state automata ("probes") whose state-to-state transitions are guarded by predicates over the values of component attributes or parameters passed by communication actions. Checking is automatically performed by the CARMA Eclipse Plugin which can generate both graphs of the transitions of the probes and graphs of the changes in underlying values within the model (such as component attributes). Using this we were able to detect from an unlabelled set of trajectories the trajectory which failed to satisfy the requirements of a specified bus route. The methods used are generally applicable to any problem where data plays a significant role and whose correctness criterion can be expressed automata-theoretically.

Our interests for future work on this topic include generating probe components directly from route descriptions which list the bus stops on the route together with their latitude and longitude coordinates. Additionally, we wish to add *instrumentation* to probes in order that they can count visits to the regions of interest on the route, enabling stronger liveness properties to be expressed.

Acknowledgements: This work is supported by the EU QUANTICOL project, 600708. We acknowledge the assistance of Transport for Edinburgh in providing access to their data on bus movement in the city of Edinburgh. Our thanks go to Natalia Zoń for her help in developing the CARMA model used in this paper. Our thanks also go to the developers of the CARMA Eclipse Plugin for providing such a useful and robust analysis platform. We are grateful to the anonymous reviewers of this paper for many helpful suggestions for improvement.

References

[1] Wil van der Aalst, Ton Weijters & Laura Maruster (2004): *Workflow mining: Discovering process models from event logs*. *Knowledge and Data Engineering, IEEE Transactions on* 16(9), pp. 1128–1142, doi:10.1109/TKDE.2004.47. Available at `http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1316839`.

[2] Yehia Abd Alrahman, Rocco De Nicola, Michele Loreti, Francesco Tiezzi & Roberto Vigo (2015): *A calculus for attribute-based communication*. In Roger L. Wainwright, Juan Manuel Corchado, Alessio Bechini & Jiman Hong, editors: *Proceedings of the 30th Annual ACM Symposium on Applied Computing, Salamanca, Spain, April 13-17, 2015*, ACM, pp. 1840–1845, doi:10.1145/2695664.2695668. Available at `http://dl.acm.org/citation.cfm?id=2695664`.

[3] Elvio Gilberto Amparore, Marco Beccuti, Susanna Donatelli & Giuliana Franceschinis (2011): *Probe Automata for Passage Time Specification*. In: *Eighth International Conference on Quantitative Evaluation of Systems, QEST 2011, Aachen, Germany, 5-8 September, 2011*, IEEE Computer Society, pp. 101–110, doi:10.1109/QEST.2011.20. Available at `http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6041098`.

[4] Elvio Gilberto Amparore & Susanna Donatelli (2010): *Model checking CSL^{TA} with Deterministic and Stochastic Petri Nets*. In: *Proceedings of the 2010 IEEE/IFIP International Conference on Dependable Systems and Networks, DSN 2010, Chicago, IL, USA, June 28 - July 1 2010*, IEEE Computer Society, pp. 605–614, doi:10.1109/DSN.2010.5544425. Available at `http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5532324`.

[5] A. Argent-Katwala, J.T. Bradley & N.J. Dingle (2004): *Expressing Performance Requirements using Regular Expressions to specify Stochastic Probes over Process Algebra Models*. In: *Proceedings of the Fourth International Workshop on Software and Performance*, ACM Press, Redwood Shores, California, USA, pp. 49–58, doi:10.1145/974044.974051.

[6] Søren Asmussen, Olle Nerman & Marita Olsson (1996): *Fitting Phase-Type Distributions via the EM Algorithm*. *Scandinavian Journal of Statistics* 23(4), pp. 419–441. Available at `http://www.jstor.org/stable/4616418`.

[7] Luca Bortolussi, Rocco De Nicola, Vashti Galpin, Stephen Gilmore, Jane Hillston, Diego Latella, Michele Loreti & Mieke Massink (2015): *CARMA: Collective Adaptive Resource-sharing Markovian Agents*. In Nathalie Bertrand & Mirco Tribastone, editors: *Proceedings Thirteenth Workshop on Quantitative Aspects of Programming Languages and Systems, QAPL 2015, London, UK, 11th-12th April 2015.*, *EPTCS* 194, pp. 16–31, doi:10.4204/EPTCS.194.2.

[8] Allan Clark & Stephen Gilmore (2008): *State-aware performance analysis with eXtended Stochastic Probes*. In Nigel Thomas & Carlos Juiz, editors: *Proceedings of the 5th European Performance Engineering Workshop (EPEW 2008)*, *LNCS* 5261, Springer, Palma de Mallorca, Spain, pp. 125–140, doi:10.1007/978-3-540-87412-6_10.

[9] Rocco De Nicola, Diego Latella, Alberto Lluch-Lafuente, Michele Loreti, Andrea Margheri, Mieke Massink, Andrea Morichetta, Rosario Pugliese, Francesco Tiezzi & Andrea Vandin (2015): *The SCEL Language: Design, Implementation, Verification*. In Martin Wirsing, Matthias M. Hölzl, Nora Koch & Philip Mayer, editors: *Software Engineering for Collective Autonomic Systems - The ASCENS Approach*, *Lecture Notes in Computer Science* 8998, Springer, pp. 3–71, doi:10.1007/978-3-319-16310-9_1.

[10] J. Hillston (1996): *A Compositional Approach to Performance Modelling*. Cambridge University Press, doi:10.1017/CBO9780511569951.

[11] Peter Kemper (2009): *Recovering model invariants from simulation traces with Petri net analysis techniques*. In: *Proceedings of the 2009 Winter Simulation Conference (WSC)*, pp. 827–838, doi:10.1109/WSC.2009.5429706. Available at `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5429706&tag=1`.

[12] Philipp Reinecke, Tilman Krauß & Katinka Wolter (2013): *Phase-Type Fitting Using HyperStar*. In Maria Simonetta Balsamo, William J. Knottenbelt & Andrea Marin, editors: *Computer Performance Engineering - 10th European Workshop, EPEW 2013, Venice, Italy, September 16-17, 2013. Proceedings*, *Lecture Notes in Computer Science* 8168, Springer, pp. 164–175, doi:10.1007/978-3-642-40725-3_13.

[13] Wil M.P. van der Aalst (2011): *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer, doi:10.1007/978-3-642-19345-3. Available at `http://link.springer.com/book/10.1007/978-3-642-19345-3`.