Population models from PEPA descriptions

Jane Hillston
LFCS, The University of Edinburgh, Edinburgh EH9 3JZ, Scotland.
Email: jeh@inf.ed.ac.uk

1 Introduction

Stochastic process algebras (e.g. PEPA [10], EMPA [1], TIPP [9]) emerged about 15 years ago as system description techniques for performance modelling. They have enjoyed some considerable success in this arena. For example, PEPA has been used to study the performance of a wide variety of systems [12, 2, 3, 18, 13].

This analysis has been based on the generation of a continuous time Markov chain (CTMC) underlying the labelled transition system of the process algebra model, derived via the interleaving structured operational semantics. The CTMC facilitates steady state and transient analysis numerically. From these distributions many performance metrics can be derived, such as utilization, throughput, and mean time to congestion. Unfortunately, as with all state-based modelling techniques, CTMCs, and consequently stochastic process algebras, suffer from problems of state space explosion. Such models can be regarded as being at the individual level as all details of all the components of the model are recorded in the state of the model and its subsequent analysis.

Although developed for performance modelling of computer and communication systems, stochastic process algebras have proved to be useful for modelling other systems as well. In particular, in recent years there has been considerable interest in using stochastic process algebras and similar formalisms for expressing models for systems biology [16, 17, 4, 6]. These models also exhibit problems of state space explosion, as both the types of components and the number of components of each type are typically large.

In recent work we have been exploring an alternative mapping from a system description in the PEPA stochastic process algebra. In this mapping we aim to capture the behaviour of PEPA components at a population level. Rather than capturing individual behaviours as happens in the CTMC semantics, we instead map to a set of non-linear ordinary differential equations (ODEs) [5, 11]. This incurs some loss of information with respect to, for example, the steady state probability distribution of the CTMC. Nevertheless the solution of the set of ODEs can still give us useful quantitative information about the system.

2 Individual vs. Population models

In most performance models the focus is on the individual entities within the model. These individuals may be customers or tasks within the system, as well as
the software and/or hardware entities of which the system is comprised. However in many modern systems the number of types of entities is not necessarily very big, while the number of entities of each type is large. This suggests that it may be better to instead view the system in terms of a small number of populations rather than individuals.

Of course making this switch to the population level results in some loss of information. It is no longer possible to precisely record the state of each entity. This information is aggregated into information about the statistics of the state of the population to which the entity belongs. However, for some performance measures at least, this population level view is sufficient. Such an abstract view has previously been adopted in the diffusion approximation technique for queueing networks [14, 15].

These alternative styles of model are already available in the systems biology arena:

**Stochastic Simulations** (Gillespie et al. [7]) are individual-based models in which each molecule is treated separately;

**Ordinary Differential Equations** (ODEs) are population-based models in which populations of molecules are represented abstractly as concentrations.

We view these not as competing approaches but as alternative analysis techniques and we would like access to both whilst also being able to carry out process algebra-based analyses such as model checking. We have modified versions of the PEPA workbench [8] which will extract underlying models amenable for three distinct forms of analysis, representing different interpretations of the PEPA system description:

- CTMC model suitable for numerical solution for steady state and transient analysis or model checking against properties expressed in the CSL modal logic;
- A representation of a CTMC suitable for simulation using Gillespie’s stochastic simulation technique;
- A set of ODEs which may be solved numerically using a variety of techniques.

The BioCHAM formalism offers a similar spectrum of interpretations and analysis techniques [6].

### 3 Deriving the ODEs

In this section we give a brief overview of how a set of ODEs may be derived from a PEPA model. The interested reader is referred to [11] for more details. There are two crucial steps. The first is to establish a more abstract representation of the state of the system than that typically used as the basis for Markovian analysis. The second is to assume that state change is a continuous rather than a discrete process.

When taking the population view of the system, rather than record the local state of each entity (current derivative) and treat the global state as the complete
set of local states, we instead record the local state of each subpopulation (the number of each possible derivative currently exhibited) and treat the global state as the set of subpopulation states. Thus we first identify the distinct entity types within the system and the local states or derivatives of each one. Each such derivative is associated with an entry in a state count vector. The number in the corresponding place in the vector records the current number of derivatives of that type.

This represents a considerably more compact state representation when there are more repeated copies of components, than there are local derivatives. However this state representation is still inherently discrete. The population view is really obtained when we make the assumption that these state variables are subject to continuous, rather than discrete, change. The evolution of each count variable can then be described by an ODE. Note that there is an implicit assumption being made here that the rate of the activity is deterministic rather than exponentially distributed.

The PEPA definitions of the component specify the activities which can increase or decrease the number of components exhibited in the current state. For any derivative the entry activities are those that bring the component into this state, while the exit activities are those enabled in this state and when performed result in a change of derivative. The cooperations show when the number of instances of another component will have an influence on the evolution of this component. This will introduce non-linearity into the set of ODEs.

Let us consider one entry of the state count vector: \( N(C_{ij}, t) \) denotes the number of \( C_{ij} \) derivatives at time \( t \).

Consider the change in a small time \( \delta t \):

\[
N(C_{ij}, t + \delta t) = N(C_{ij}, t) - \sum_{(a,r) \in Ex(C_{ij})} r \times \min_{c_{ij} \in Ex(a,r)} (N(C_{ij}, t)) \delta t + \sum_{(a,r) \in En(C_{ij})} r \times \min_{c_{ij} \in Ex(a,r)} (N(C_{ij}, t)) \delta t
\]

Dividing by \( \delta t \) and taking the limit, \( \delta t \rightarrow 0 \) we obtain:

\[
\frac{dN(C_{ij}, t)}{dt} = - \sum_{(a,r) \in Ex(C_{ij})} r \times \min_{c_{ij} \in Ex(a,r)} (N(C_{ij}, t)) + \sum_{(a,r) \in En(C_{ij})} r \times \min_{c_{ij} \in Ex(a,r)} (N(C_{ij}, t))
\]

In practice we do not wish to derive the system of ODEs representing a PEPA model from first principles in this way. Instead the derivation proceeds...
4 Conclusions

The use of stochastic process algebra models for Markovian based modelling is well-established, but hampered by the problem of state space explosion. In some models the problems arise because there are many instances of components of the same type. In this case the use of more abstract mathematical models which focus on the statistics of the population rather than the individuals seems promising. We have recently established a mapping to such a mathematical model for the stochastic process algebra PEPA [11]. This is an alternative, not a replacement for existing techniques.

We are currently exploring the relationship between models derived in this way and other models which can be derived from the PEPA description. Our PEPA models of biochemical pathways have been experimenting with different levels of abstraction. Whereas treating each molecule in the system as a distinct entity or component in the model corresponds to the stochastic simulation view of the system, we have developed models with discrete levels of concentration, corresponding to the ODE view [4]. These models can also be interpreted as CTMCs and we are investigating the relationship between the ODE and the CTMC obtained in these cases (see Figure 1).
References
