# Coping with the Parallelism of BitTorrent: Conversion of PEPA to ODEs in Dealing with State Space Explosion

Adam Duguid

Laboratory of Computer Science, The University of Edinburgh
a.j.duguid@sms.ed.ac.uk

**Abstract.** The Performance Evaluation Process Algebra (PEPA) language is a stochastic process algebra, generating Continuous Time Markov Chains (CTMC) to allow quantitative analysis. Protocols such as BitTorrent are highly parallel in nature, and represent one area where CTMC analysis is limited by the well-known state space problem. The number of unique states each client can exist in, and the number of clients required to accurately model a typical BitTorrent network preclude the use of CTMCs. Recent work has shown that PEPA models also allow the derivation of an activity matrix, from which ODE and stochastic simulation models, as alternative forms of analysis, are possible. Using this technique, a BitTorrent network is created, analysed, and the results compared against previous BitTorrent models.

## 1 Introduction

The PEPA[1] language originated, in part, from Calculus of Communicating Systems (CCS), allowing the generation of a labelled transition graph with rates based on the exponential distribution. From this graph a CTMC can be obtained and the steady state gained through standard numerical techniques. CTMCs produce exact results, in the sense that every possible state of the system is accounted for and that all probabilities are correct for the given rates. The use of CTMCs assumes that the subsystems behaviour can, to some degree of accuracy, be described with exponential distributions and behaviour is independent of time.

One particular weakness of CTMCs is the size of the model which can be efficiently analysed while still remaining tractable. As the number of components increases, especially components that act independently from one another, the size of the state space can rapidly expand beyond tractable limits. Techniques such as model simplification and state aggregation can allow the analysis of larger models to some extent but the limitations still remain.

Recent work has introduced process algebra (in this instance $\pi$ calculus) to the area of systems biology[2,3], a field interested in the dynamic pathways of biological systems. However, with the desire to model large numbers of proteins or receptors, the dominant approach within systems biology has been to use

E. Asarin and P. Bouyer (Eds.): FORMATS 2006, LNCS 4202, pp. 156–170, 2006.

Ordinary Differential Equations (ODEs). ODEs are continuous-time, continuous-state and deterministic in nature. In addition, the area of ODEs is well researched and so supported by a range of solvers. The modelling of biological systems can also be conducted through stochastic simulation such as Gillespie's Stochastic Simulation Algorithm (SSA)[4]. Gillespie's argument for the use of a continuous-time, discrete-state stochastic simulation centres on physical accuracy with the real system, in this case chemical reactions. Both approaches scale differently to CTMCs, where the numbers of each component do not affect the complexity of the model in the same way.

This has led to work mapping PEPA to ODEs[5,6] to model the Extracellular signal Regulated Kinase (ERK) signalling pathway from within a process algebra. Two structures were defined, reagent-centric and pathway-centric, offering different views of the system. Both structures allow the derivation of ODEs from the underlying PEPA model, offering an alternative style of analysis from CTMCs. This same derivation process can be used to allow the use of SSA.

The PEPA language will be briefly covered, followed by a more complete description of the mapping from PEPA to ODEs. The salient parts of the BitTorrent protocol will be detailed, and a PEPA model constructed using the reagent-centric approach. The BitTorrent protocol is used due to its parallel nature of communication between entities, which resists analysis with CTMCs. Attempting to model even 20 peers can easily lead to a state space as large as $100^{20}$ with CTMCs while remaining tractable when using ODEs or SSAs. Finally, these results will be compared against an existing model of BitTorrent.

## 2 PEPA

A model defined in the PEPA language consists of a number of components representing different agents or entities in the real system. The components interact with each other through a small set of combinators as shown below.

$$P ::= (\alpha, r).P \mid P + Q \mid P \underset{L}{\bowtie} Q \mid P/L \mid A$$

**Prefix** $(\alpha, r).P$ represents a component that can perform an activity $\alpha$ at rate $r$ (sampled from the negative exponential distribution) before it transitions to a component of type $P$.

**Choice** $P + Q$ represents a component which is either of type $P$ or $Q$. Which is chosen is based on a race condition on the first activity of each component.

**Co-operation** $P \underset{L}{\bowtie} Q$ requires that if components $P$ and $Q$ can both perform an activity $\alpha$ (where $\alpha \in L$), then for either component to perform $\alpha$, they must both perform it together. Where P or Q are capable of an activity not in the set L then these can occur independently.