

Approximate reasoning for real-time probabilistic processes

Vineet Gupta
Google Inc.
vineet@google.com

Radha Jagadeesan*
School of CTI, DePaul University
rjagadeesan@cs.depaul.edu

Prakash Panangaden†
School of Computer Science, McGill University
prakash@cs.mcgill.ca

Abstract

We develop a pseudo-metric analogue of bisimulation for generalized semi-Markov processes. The kernel of this pseudo-metric corresponds to bisimulation; thus we have extended bisimulation for continuous-time probabilistic processes to a much broader class of distributions than exponential distributions. This pseudo-metric gives a useful handle on approximate reasoning in the presence of numerical information - such as probabilities and time - in the model.

We give a fixed point characterization of the pseudo-metric. This makes available coinductive reasoning principles for reasoning about distances. We demonstrate that our approach is insensitive to potentially ad hoc articulations of distance by showing that it is intrinsic to an underlying uniformity. We provide a logical characterization of this uniformity using a real-valued modal logic.

We show that several quantitative properties of interest are continuous with respect to the pseudo-metric. Thus, if two processes are metrically close, then observable quantitative properties of interest are indeed close.

1. Introduction

The starting point and conceptual basis for classical investigations in concurrency are the notions of equivalence and congruence of processes — when can two processes be considered the same and when can they be substituted for each other. Most investigations into

timed [1, 2] and probabilistic concurrent processes are based on equivalences of one kind or another, e.g. [15, 18, 20, 25, 9, 23, 22] to name but a few.

As has been argued before [19, 11, 13], this style of reasoning is fragile in the sense of being too dependent on the exact numerical values of times and probabilities. Previously this had pointed out for probability, but the same remarks apply, *mutis mutandis*, to real time as well. Consider the following two paradigmatic examples:

- Consider the probabilistic choice operator: $\mathcal{A}_1 +_p \mathcal{A}_2$, which starts \mathcal{A}_1 with probability p and \mathcal{A}_2 with probability $1 - p$. Consider $\mathcal{A}_1 +_{p+\epsilon} \mathcal{A}_2$ and $\mathcal{A}_1 +_{p+2\epsilon} \mathcal{A}_2$. In classical exact reasoning, the best that one can do is to say that all these three processes are inequivalent. Clearly, there is a gradation here: $\mathcal{A}_1 +_{p+\epsilon} \mathcal{A}_2$ is closer to $\mathcal{A}_1 +_p \mathcal{A}_2$ than $\mathcal{A}_1 +_{p+2\epsilon} \mathcal{A}_2$ is to $\mathcal{A}_1 +_p \mathcal{A}_2$.
- Consider the $\text{delay}_t.\mathcal{A}$ operator that starts \mathcal{A} after a delay of t time units. Consider $\text{delay}_{t+\epsilon}.\mathcal{A}$ and $\text{delay}_{t+2\epsilon}.\mathcal{A}$. Again, in exact reasoning, the best that one can do is to say that all these three processes are inequivalent. Again, $\text{delay}_{t+\epsilon}.\mathcal{A}$ is intuitively closer to $\text{delay}_t.\mathcal{A}$ than $\text{delay}_{t+2\epsilon}.\mathcal{A}$ is to $\text{delay}_t.\mathcal{A}$.

In both examples, the intuitive reasoning behind relative distances is supported by calculated numerical values of quantitative observables — expectations in the probabilistic case and (cumulative) rewards in the timed case.

The fragility of exact equivalence is particularly unfortunate for two reasons: firstly, the timings and probabilities appearing in models should be viewed as numbers with some error estimate. Secondly, probability distributions over uncountably many states arise in even superficially discrete paradigms such as Generalized semi-

* Research supported in part by NSF 0244901

† Research supported in part by NSERC and EPSRC.

markov processes (e.g. see [24] for a textbook survey), and discrete approximations are used for algorithmic purposes [4, 16]. These approximants do not match the continuous state model exactly and force us to think about approximate reasoning principles — e.g. when does it suffice to prove a property about an approximant?

Thus, we really want an “approximate” notion of equality of processes. In the probabilistic context, Jou and Smolka [19] propose that the correct formulation of the “nearness” notion is via a metric. Similar reasons motivate the study of Lincoln, Mitchell, Mitchell and Scedrov [21], our previous study of metrics for labelled Markov processes [11, 13, 12], the study of the fine structure of these metrics by van Breugel and Worrell [8, 7] and the study of Alfaro, Henzinger and Majumdar of metrics for probabilistic games [10].

In contrast to these papers, in the present paper we focus on real-time probabilistic systems that combine *continuous time and probability*. We consider generalized semi-Markov processes (GSMP). GSMPs strictly generalize continuous time Markov chains by permitting general (i.e. non exponential) probability distributions.

Following the format of the usual definition of bisimulation as a maximum fixed point, we define a metric on states of a GSMP as a maximum fixed point. This permits us to use traditional styles of reasoning to reason about metric distances. For example, in exact reasoning, to deduce that two states are equivalent, it suffices to produce a bisimulation that equates the states. In our setting, to show that the distance between two states is less than ϵ , it suffices to produce a (metric) bisimulation that sets the distance between the states to be $< \epsilon$.

Viewing metric distance 0 as bisimilarity, we get a definition of bisimulation for GSMPs, a class that properly includes CTMCs. In contrast to existing work on bisimulation for general probability distributions (e.g. [6, 17]) our definition accounts explicitly for the change of probability densities over time.

Secondly, we demonstrate that our study does not rely on “ad-hoc” construction of metric distances. Uniform spaces capture the essential aspects of metric distances by axiomatizing the structure needed to capture relative distances — e.g. statements of the form “x is closer to y than to z”. A metric determines a uniform space but different metrics can yield the same uniform space. Uniform spaces represent more information than topological spaces but less than metric spaces, so we are identifying, as precisely as we can, the intrinsic source of the quantitative information. We present our maximal fix-point construction as a construction on uniform spaces, showing that the numerical values of different metric

representations are not used in an essential way. In particular, in our setting, it shows that the actual numerical values of the discount factors used in the definition of the metric do not play any essential role.

Thirdly, we provide a “logical” characterization of the uniformity using a real-valued modal logic. In analogy to traditional completeness results, we prove that the uniformity notion induced by the real-valued modal logic coincides with the uniformity induced by the metric defined earlier. Our logic is intentionally chosen to prove this completeness result. It is not intended to be used as an expressive specification formalism to describe properties of interest. Our framework provides an intrinsic characterization of the quantitative observables that can be accommodated — functions that are continuous wrt the metric.

Finally, we illustrate the use of such studies in reasoning by showing that several quantitative properties of interest are continuous with respect to the metric. Thus, if two processes are close in the metric then observable quantitative properties of interest are indeed close. For expository purposes, the list considered in this paper includes expected hitting time, expected (cumulative and average) rewards. The tools used to establish these results are “continuous mapping theorems” from stochastic process theory, and provide a general recipe to tackle other observables of interest.

Organization of paper.

The rest of this paper is organized as follows. We begin with a review of the model of GSMPs in Section 2. We then give a review of the basic ideas from stochastic process theory — metrics on probability measures on metric spaces in Section 4 and the Skorohod J2 metrics on timed traces in Section 5. We define metric bisimulation in Section 7. We represent our construction in terms of uniform spaces in Section 8. We discuss interesting quantitative observables are continuous functions in Section 9, and. We show the completeness of the real-valued modal logic in Section 10.

2. Generalized semi-Markov processes

GSMPs properly include finite state CTMCs while also permitting general probability distributions. We describe GSMPs informally here following the formal description of [27]. The key point is that in each state there are possibly several events that can be executed. Each event has its own clock - running down at its own rate - and when the first one reaches zero that event is selected

for execution. Then a probabilistic transition determines the final state and any new clocks are set according to given probability distributions: defined by conditional density functions.

A finite-state GSMP over a set of propositions AP and has the following ingredients:

1. A finite set S of states. Each state has an associated finite set of events $I(s)$, each with its own clock (we use the same letter for an event and its clock) and a non-zero rate for each clock in $I(s)$. A clock in $I(s)$ runs down at the constant rate associated with it.
2. A labelling function $\text{PROPS} : S \rightarrow 2^{AP}$ that assigns truth values to atomic propositions in each state.
3. A continuous probability density function $f(t; s, i; s', i')$, over time, for each $i \in I(s)$, for each target state s' and $i' \in I(s')$.
4. For each $i \in I(s)$, a probabilistic transition function $\text{next}_i : S \times S \rightarrow [0, 1]$. We require $\sum_{s' \in S} \text{next}_i(s, s') = 1$.

We use \tilde{c}, \tilde{c}' (resp. \tilde{r}) for vectors of clock-values (resp. rates). We use the vector operation $\tilde{c} - \tilde{r}t$ to indicate the clock vector resulting from evolution of each clock under its rate for time t .

Definition 2.1 *Let s be a state. A generalized state is of the form $\langle s, \tilde{c} \rangle$ where c is a vector of clock-values indexed by $i \in I(s)$ that satisfies a uniqueness condition: there is a unique clock in $I(s)$ that reaches 0 first.*

We write $\mathcal{T}(\langle s, \tilde{c} \rangle)$ for the time required for the first clock (unique by the above definition) to reach 0. We use \mathcal{G} for the set of generalized states, and $\mathbf{g}_s, \mathbf{g}'_s, \mathbf{g}_s^1 \dots$ for generalized states.

We describe the evolution starting in a generalized state $\langle s, \tilde{c} \rangle$. Each clock in \tilde{c} decreases at its associated rate. By uniqueness condition on generalized states, a unique clock reaches 0 first. Let this clock be $i \in I(s)$. The distribution on the next states is determined by the probabilistic transition function $\text{next}_i : S \times S \rightarrow [0, 1]$. For each target state s' ,

- The clocks $i' \in I(s) \setminus I(s')$ are discarded.
- The new clocks $i' \in I(s') \setminus [I(s) \setminus \{i\}]$, get new initial timevalues assigned as per the continuous probability density function $f(s, i; s', i')$.
- The remaining clocks carry forward their time values from s to s'

The continuity condition on probability distributions ensures that this informal description yields a legitimate Markov kernel [27]. The semantics of a real-time probabilistic process can be described as a discrete-time Markov process on generalized states. For each generalized state, we associate a set of sequences of generalized states that arise following the prescription of evolution given above. After reviewing some metric space and stochastic process theory we return to GSMPs.

3. Pseudometrics

Definition 3.1 *A pseudometric m on a state space S is a function $S \times S \rightarrow [0, 1]$ such that:*

$$d(x, x) = 0, d(x, y) = d(y, x), d(x, z) \leq d(x, y) + d(x, z)$$

A function $f : (M, m) \rightarrow (M', m')$ is Lipschitz if $(\forall x, y) m'(f(x), f(y)) \leq m(x, y)$.

We consider a partial order on pseudometrics on a fixed set of states S .

Definition 3.2 *\mathcal{M} is the class of pseudometrics on S ordered as: $m_1 \preceq m_2$ if $(\forall s, t) m_1(s, t) \geq m_2(s, t)$.*

Lemma 3.3 *(\mathcal{M}, \preceq) is a complete lattice.*

The top element \top is the constant 0 function, and the bottom element is the discrete metric [12]. Thus, any monotone function F on (\mathcal{M}, \preceq) has a complete lattice of fixed points.

4. Wasserstein metric

Let (M, m) be a pseudometric space, and let P, Q be probability measures on M . Then, $W(m)(P, Q)$ is defined by the solution to the following linear program ($h : M \rightarrow [0, 1]$ is any function):

$$\begin{aligned} W(m)(P, Q) = & \sup_h \int h dP - \int h dQ \\ \text{subject to : } & \forall s \in M. h(s) \leq 1 \\ & \forall s, s'. |h(s) - h(s')| \leq m(s, s'). \end{aligned}$$

An elementary exercise using the linear program shows that the distances on distributions satisfies symmetry and the triangle inequality, so we get a pseudometric, written $W(m)$, on distributions.

By standard results (see Anderson and Nash [3]), this is equivalent to defining $W(m)(P, Q)$ as the solution to the following dual linear program (here ρ is any measure

on $M \times M$, S is any P -measurable subset and S' is any Q -measurable subset.):

$$W(m)(P, Q) = \inf \int m \, d\rho$$

subject to :

$$\begin{aligned} \forall S, \rho. \rho(S \times M) &= P(S) \\ \forall S', \rho. \rho(M \times S') &= Q(S') \\ \forall S, S'. \rho(S \times S') &\geq 0. \end{aligned}$$

The Wasserstein construction is monotone on the lattice of pseudometrics.

Lemma 4.1 $m \preceq m' \Rightarrow W(m) \preceq W(m')$

We discuss some concrete examples to illustrate the distances yielded by this construction. Let (M, m) be a 1-bounded metric space, ie. $m(x, y) \leq 1$, for all x, y . Let 1_x be the unit measure concentrated at x ,

Example 4.2 We calculate $W(m)(1_x, 1_{x'})$. Consider the function $f : M \rightarrow [0, 1]$ by $f(y) = m(x, y)$. Using the primal linear program yields that $W(m)(1_x, 1_{x'}) \leq m(x, x')$. On the other hand using the dual linear program yields $W(m)(1_x, 1_{x'}) \geq m(x, x')$.

Example 4.3 Let P, Q be such that for all measurable U , $|P(U) - Q(U)| < \epsilon$. For any 1-bounded function h , $|\int h dP - \int h dQ| < \epsilon$ — since for any simple function g with finite range $\{v_1 \dots v_n\}$ dominated by h :

$$\begin{aligned} &\sum_i v_i P(g^{-1}(v_i)) - v_i Q(g^{-1}(v_i)) \\ &\leq \sum_i v_i [P(g^{-1}(v_i)) - Q(g^{-1}(v_i))] \\ &\leq \sum_i P(g^{-1}(v_i)) - Q(g^{-1}(v_i)) \\ &= P(g^{-1}(\{v_1 \dots v_n\})) - Q(g^{-1}(\{v_1 \dots v_n\})) \\ &< \epsilon \end{aligned}$$

So, $W(m)(P, Q) < \epsilon$.

The following lemma is the key tool to approximate continuous probability distributions by discrete distributions in a separable metric space. We use the following lemma later with the $U_i, i > 1$'s being subsets of ϵ -neighborhoods of a point, and the $U_i, i > 1$ being a finite cover, wrt P , for all but ϵ of the space.

Lemma 4.4 Let P be a probability measure on (M, m) . Let $U_i, i = 0, 1, 2, \dots, n$ be a finite partition of the points of M into measurable sets such that:

- $(\forall i \geq 1) [\text{diameter}(U_i) \leq \epsilon]$ ¹
- $P(U_0) \leq \epsilon$

Let x_i be such that $x_i \in U_i$. Define a discrete probability measure Q on (M, m) by: $Q(\{x_i\}) = P(U_i)$. Then:

$$W(m)(P, Q) \leq 2\epsilon$$

¹ The diameter of a set S is $\sup\{m(x, y) \mid x, y \in S\}$

5. Cadlag functions

Definition 5.1 Let (M, m) be a pseudometric space. $f : [0, \infty) \rightarrow M$ is cadlag if for any decreasing sequence $\{t\} \downarrow t_0$

$$\lim_{t \rightarrow t_0} f(t) = f(t_0)$$

and for any increasing sequence $\{t\} \uparrow t_0$

$$\lim_{t \rightarrow t_0} f(t) \text{ exists}$$

We write $\mathcal{D}^{(M, m)}[0, \infty)$ (or $\mathcal{D}^m[0, \infty)$, when M is clear from context) for cadlag functions with range (M, m) .

Let (M, m) be a metric space. Let $|\cdot|$ be the metric on positive reals R^+ be defined by

$$|\cdot|(r, r') = |r - r'|$$

Definition 5.2 (Skorohod J2 metric)

Let (M, m) be a metric space. Let f, g be cadlag functions: $[0, \infty) \rightarrow M$. $J(m)(f, g)$ is defined as:

$$\max\left(\sup_t \inf_{t'} [\max(m(f(t), g(t')), |t - t'|)], \sup_{t'} \inf_t [\max(m(f(t), g(t')), |t - t'|)] \right)$$

Thus the J2 distance between two functions is the Hausdorff distance between their graphs (ie. the set of points $(x, f(x))$ in the space $[0, \infty) \times M$ equipped with the metric $d((x, s), (y, t)) = \max(|x - y|, m(s, t))$.

The following lemma is immediate from definitions.

Lemma 5.3 $m_1 \preceq m_2 \Rightarrow J(m_1) \preceq J(m_2)$

The next lemma is standard, e.g. see Billingsley [5].

Lemma 5.4 (Skorohod)

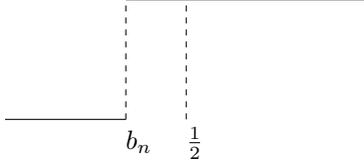
If (M, m) is separable, $\mathcal{D}^M[0, \infty)$ is a separable space with a countable basis given by piecewise constant functions with finitely many discontinuities and finite range contained in a basis of M .

We consider a few examples, to illustrate the metric — see the book by Whitt [26] for a detailed analysis of Skorohod's metrics. The first example shows that jumps/discontinuities can be matched by closeby jumps.

Example 5.5 [26] Let $\{b_n\}$ be an increasing sequence that converges to $\frac{1}{2}$. Consider:

$$f_{b_n}(r) = \begin{cases} 0, & r < b_n \\ 1, & r \geq b_n \end{cases}$$

These are depicted in the picture below.



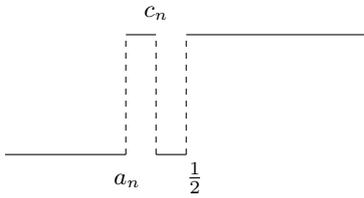
The sequence $\{f_{b_n}\}$ converges to $f_{\frac{1}{2}}$.

The next example shows that a single jumps/discontinuities can be matched by multiple closeby jumps.

Example 5.6 [26] Let $\{a_n\}, \{c_n\}$ be increasing sequences that converges to $\frac{1}{2}$ such that $a_n < c_n$. Let:

$$g_n(r) = \begin{cases} 0, & r < a_n \\ 1, & a_n \leq r < c_n \\ 0, & c_n \leq r < \frac{1}{2} \\ 1, & r \geq \frac{1}{2} \end{cases}$$

These are depicted in the picture below.



The sequence $\{g_n\}$ converges to $f_{\frac{1}{2}}$.

The next two non-examples shows that “jumps are detected”. In a later section, we develop a real-valued modal logic that captures the reasoning behind these two non-examples. Here, to provide preliminary intuitions, we provide a preview of this development in a specialized form. Given a cadlag function with range $[0, 1]$, and a Lipschitz operator h on $[0, 1]$ let $\mathcal{L}(h)(f)(t)$ be the upper Lipschitz approximation to $h \circ f$ evaluated at t , i.e.

$$\mathcal{L}(h)(f)(t) = \sup\{h(f(t')) - |t' - t| \mid t' \in [0, \infty)\}$$

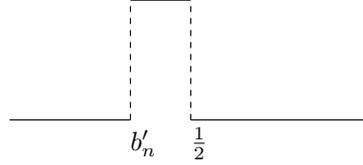
In this definition, view h as a test performed on the values taken by the function f — since h is a Lipschitz operator on $[0, 1]$, the results of such tests are smooth, and include the analogue of (logical) negation via the operation $1 - (\cdot)$ and smooth conditionals via $h_q(x) = \max(0, x - q)$ that correspond to a “greater than q ” test. The $\mathcal{L}(h)(f)$ also performs an extra smoothing operation over time, so that the values of $\mathcal{L}(h)(f)$ at times t, t' differ by atmost $|t - t'|$.

The first non-example shows that jumps are detected — a sequence of functions with jumps cannot converge to a continuous function.

Example 5.7 Let $\{b'_n\}$ be an increasing sequence that converges to $\frac{1}{2}$. Consider:

$$f_{b'_n}(r) = \begin{cases} 0, & r < b'_n \\ 1, & b'_n \leq r < \frac{1}{2} \\ 0, & \frac{1}{2} \leq r \end{cases}$$

These are depicted in the picture below.



The sequence $\{f_{b'_n}\}$ does not converge to the constant 0 function $\mathbf{0}$. This is illustrated by considering the Lipschitz operator h'_ϵ on $[0, 1]$ defined as:

$$h'_\epsilon(r) = \max(0, \epsilon - |1 - r|)$$

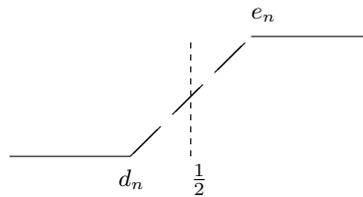
For all n , $\mathcal{L}(h)(g_n)(\frac{1}{2}) = \epsilon$, but $\mathcal{L}(h)(\mathbf{0})(\frac{1}{2}) = 0$.

The second non-example shows that continuous functions do not approximate a function with jumps.

Example 5.8 [26] Let $\{\epsilon_n\}$ be an decreasing sequences that converges to 0. Let $d_n = \frac{1}{2} - \frac{\epsilon_n}{2}$, $e_n = \frac{1}{2} + \frac{\epsilon_n}{2}$. Consider:

$$h_n(r) = \begin{cases} 0, & r < d_n \\ \frac{r-d_n}{\epsilon_n}, & d_n \leq r < e_n \\ 1, & r \geq e_n \end{cases}$$

These are depicted in the picture below.



The sequence $\{g_n\}$ does not converge to $f_{\frac{1}{2}}$. To analyze in terms of the operator \mathcal{L} , consider the Lipschitz operator h_ϵ on $[0, 1]$ defined as:

$$h_\epsilon(r) = \max(0, \epsilon - |\frac{1}{2} - r|)$$

For all n , $\mathcal{L}(h)(g_n)(\frac{1}{2}) = \epsilon$, but $\mathcal{L}(h)(f_{\frac{1}{2}})(\frac{1}{2}) = 0$.

We conclude this section with a discussion of a delay operator on the space of cadlag functions.

Definition 5.9 Let (M, m) be a metric space. Let $f \in \mathcal{D}^{(M, m)}[0, \infty)$. Let $s \in M, 0 \leq r$. Let $u : [0, r) \rightarrow M$ be a continuous function. Define $\text{delay}_u(f) \in \mathcal{D}^{(M, m)}[0, \infty)$ as follows:

$$\text{delay}_u(f)(t) = \begin{cases} f(t-r), & t \geq r \\ u(t), & 0 \leq t < r \end{cases}$$

The distance between a cadlag function and its t -delayed version is no greater than t .

Lemma 5.10 Let $0 \leq r$. Let $u : [0, r) \rightarrow M$ be continuous such that $(\forall 0 \leq r' < r) m(u(r'), f(0)) \leq r$. Then: $J(m)(\text{delay}_u(f), f) \leq r$.

6. GSMPs and cadlag functions.

We deal with the temporal aspects of GSMPs next by constructing cadlag functions for paths (sequences of generalized states of a GSMP).

A sequence of generalized states is *finitely varying* if it is non-Zeno, i.e. for any i , $\sum_{j>i}^{\infty} \mathcal{T}(\langle s_j, \tilde{\mathbf{c}}_j \rangle)$ diverges. Any finitely varying sequence of generalized states $\langle s_i, \tilde{\mathbf{c}}_i \rangle$ generates $f : [0, \infty) \rightarrow \mathcal{G}$ as follows:

- $f(t) = \langle s_i, \tilde{\mathbf{c}} \rangle$, where

$$\sum_{k=0}^i \mathcal{T}(\langle s_k, \tilde{\mathbf{c}}_k \rangle) \leq t < \sum_{k=0}^{i+1} \mathcal{T}(\langle s_k, \tilde{\mathbf{c}}_k \rangle)$$

$$\text{and } \tilde{\mathbf{c}} = \tilde{\mathbf{c}}_i - \tilde{\mathbf{r}}_{\mathbf{c}} | t - \sum_{k=0}^i \mathcal{T}(\langle s_k, \tilde{\mathbf{c}}_k \rangle) |$$

Such finitely varying traces satisfy for any interval $[t, t']$, that there is a finite partition $t = t_0 < t_1 < t_2 \dots < t_n = t'$ such that:

- If $f(t_i) = \langle s, \tilde{\mathbf{c}} \rangle$, then: $(\forall t_i \leq t < t_{i+1}) f(t) = \langle s, \tilde{\mathbf{c}} - \tilde{\mathbf{r}}_{\mathbf{c}} | t - t_i \rangle$.

We write $\text{Traces} \langle s, \tilde{\mathbf{c}} \rangle$ for the set of traces that start with $\langle s, \tilde{\mathbf{c}} \rangle$. The probability distributions associated with initial clock-values at states (f_s) and transitions (next) induces a probability measure on $\text{Traces} \langle s, \tilde{\mathbf{c}} \rangle$. The paths that are zeno have measure zero, so the finitely-varying paths generate $\text{Traces} \langle s, \tilde{\mathbf{c}} \rangle$ in measure.

Arbitrarily close approximations to the distance between finitely-varying functions $f, g \in \mathcal{D}^{(\mathcal{G}, m)}[0, \infty)$ are forced by the distances between the values of f, g at finitely many points of time. This lemma is useful later on to show that our coinductive definition of metric has closure ordinal ω .

Lemma 6.1 Let m be a pseudometric over generalized states. Let $f, g \in \mathcal{D}^{(\mathcal{G}, m)}[0, \infty)$ be finitely-varying functions such that $J(m)(f, g) > \delta$.

Then there is a finite subset $\mathcal{G}_f \subseteq \mathcal{G}$ and $\epsilon > 0$ such that for any $m', J(m')(f, g) > \delta$ if $(\forall \mathbf{g}_s, \mathbf{g}'_s \in \mathcal{G}_f) [m'(\mathbf{g}_s, \mathbf{g}'_s) \geq m(\mathbf{g}_s, \mathbf{g}'_s) - \epsilon]$.

7. Bisimulation style definition of metric

Let \mathcal{M} be the class of pseudo-metrics on generalized states that satisfy:

$$\text{Props}(s) \neq \text{Props}(s') \Rightarrow m(\langle s, \tilde{\mathbf{c}} \rangle, \langle s', \tilde{\mathbf{c}}' \rangle) = 1$$

and ordered as in section 3: $m_1 \preceq m_2$ if $(\forall s, t) m_1(s, t) \geq m_2(s, t)$. Fix $0 < k < 1$. Define a functional \mathcal{F}_k on \mathcal{M} :

Definition 7.1 $\mathcal{F}_k(m)(\langle s, \tilde{\mathbf{c}} \rangle, \langle s', \tilde{\mathbf{c}}' \rangle) < \epsilon$ if

$$k \times W(J(m))(\text{Traces} \langle s, \tilde{\mathbf{c}} \rangle, \text{Traces} \langle s', \tilde{\mathbf{c}}' \rangle) < \epsilon$$

In this definition, view k as a discount factor. In the next section, we will show that the choice of k doesn't affect the essential character of the metric.

As an immediate consequence of lemmas 5.3 and 4.1:

Lemma 7.2 \mathcal{F}_k is monotone on \mathcal{M} .

Since (\mathcal{M}, \preceq) is a complete lattice, using Tarski's fixed point theorem, \mathcal{F}_k has a maximum fixed point, $m_{\mathcal{F}_k}$.

Definition 7.3 m is a metric-bisimulation if $m \preceq \mathcal{F}_k(m)$.

By standard results:

$$m_{\mathcal{F}_k} = \bigsqcup \{m \mid m \text{ is a metric-bisimulation}\}$$

Thus, a metric-bisimulation m provides an upper bound on the distances assigned by $m_{\mathcal{F}_k}$.

Example 7.4 Consider the equivalence relation $\simeq = \{(\langle s, \tilde{\mathbf{c}} \rangle, \langle s', \tilde{\mathbf{c}}' \rangle) \mid m_{\mathcal{F}_k}(\langle s, \tilde{\mathbf{c}} \rangle, \langle s', \tilde{\mathbf{c}}' \rangle) = 0\}$. \simeq describes a notion of bisimulation and is explicitly defined as follows.

Let $\mathcal{M}_{\{0,1\}}$, sublattice of \mathcal{M} consisting of metrics whose range is $\{0, 1\}$, i.e. all distances are either 0 or 1. $\mathcal{M}_{\{0,1\}}$ is essentially the class of equivalence relations. A simple proof shows that:

$$\simeq = \bigcup \{m^{-1}(0) \mid m \in \mathcal{M}_{\{0,1\}}, m \preceq \mathcal{F}_k(m)\}$$

As an example of metric-reasoning, we now show that generalized states with the same state, but clock values reflecting evolution for a time t are m_F -close.

Lemma 7.5 Define a pseudometric m on generalized states as follows:

$$m(\langle s, \tilde{c} \rangle, \langle s', \tilde{c}' \rangle) = \begin{cases} \min(1, t), & \text{if } s = s' \text{ and} \\ & \tilde{c} = \tilde{c}' + \tilde{\mathbf{r}}_c t \text{ or} \\ & \tilde{c} = \tilde{c}' - \tilde{\mathbf{r}}_c t \\ 1, & \text{otherwise} \end{cases}$$

Then: $m \preceq \mathcal{F}_k(m)$.

The role played by the discount constant k is captured in the following fact: $(\forall \langle s, \tilde{c} \rangle, \langle s', \tilde{c}' \rangle) |m_{n+1}(\langle s, \tilde{c} \rangle, \langle s', \tilde{c}' \rangle) - m_n(\langle s, \tilde{c} \rangle, \langle s', \tilde{c}' \rangle)| \leq k^{n+1}$. This is the key step in the proof of the following lemma.

Lemma 7.6 $m_{\mathcal{F}_k}$ is separable.

The separability of $m_{\mathcal{F}_k}$ enables to prove the analogue of lemma 6.1.

Lemma 7.7 (Finite detectability of distances)

Let m be a pseudometric on \mathcal{G} with countable basis. Let $W(J(m))(\langle s, \tilde{c} \rangle, \langle s', \tilde{c}' \rangle) > \delta$.

Then there is a finite subset $\mathcal{G}_f \subseteq \mathcal{G}$ and $\epsilon > 0$ such that for any m' , $W(J(m'))(\langle s, \tilde{c} \rangle, \langle s', \tilde{c}' \rangle) > \delta$ if $(\forall \mathbf{g}_s, \mathbf{g}'_s \in \mathcal{G}_f) [m'(\mathbf{g}_s, \mathbf{g}'_s) \geq m(\mathbf{g}_s, \mathbf{g}'_s) - \epsilon]$.

Lemma 7.8 \mathcal{F}_k has closure ordinal ω .

Proof. The proof proceeds by showing that the maximum fixed point m is given by $m = \sqcup_i m_i$, where $m_0 = \top$ and $m_{i+1} = \mathcal{F}_k(m_i)$.

Let $m(\langle s, \tilde{c} \rangle, \langle s', \tilde{c}' \rangle) > \epsilon$. From lemma 7.7, we deduce the finitely many conditions of the form $m'(\langle s_i, \tilde{c}_i \rangle, \langle s'_i, \tilde{c}'_i \rangle) > \epsilon_i$ that suffice to ensure that $F(m') > \epsilon$. Each of these finitely many conditions are met at a finite index. Result follows. ■

8. Uniform spaces

A uniform space — e.g. see Geroch [14] for a quick survey — captures the essence of the distance notion in metric spaces, e.g. if there are points x, y, z such that x is closer to y than to z , uniform spaces have enough data to capture this without committing to the actual numerical values of the distances. The aim of this section is to show that our treatment is “upto uniformity” — this is a formal way of showing that there is no ad-hoc treatment of the quantitative metric distances. In particular, we show that different discount factors k yield the same uniformity.

Let \mathcal{S} be a set of states.

Definition 8.1 A pseudo-uniformity, \mathcal{U} is a collection of subsets of $\mathcal{S} \times \mathcal{S}$, called entourages, that satisfies:

- $(\forall E \in \mathcal{U}) (\forall x \in \mathcal{S}) (x, x) \in E$
- $E \in \mathcal{U} \Rightarrow E^{-1} \in \mathcal{U}$
- $E \in \mathcal{U} \Rightarrow (\exists E' \in \mathcal{U}) E'E' \subseteq E$
- $E, E' \in \mathcal{U} \Rightarrow E \cap E' \in \mathcal{U}$
- $E \in \mathcal{U}, E \subseteq E' \Rightarrow E' \in \mathcal{U}$

The usual presentation of uniformities also includes

$$\bigcap_{E \in \mathcal{U}} E = \{\langle x, x \rangle\}$$

This condition is not appropriate to the pseudo-metric setting that we are in. To gain intuition into this definition, we describe how a pseudometric generates a pseudo-uniformity. Given a pseudo-metric m on \mathcal{S} , let $K_m^\epsilon = \{(x, y) \mid m(x, y) < \epsilon\}$. We get a pseudo-uniformity by considering:

$$\{E \mid (\exists \epsilon)\{(x, y) \mid m(x, y) < \epsilon\} \subseteq E\}$$

Thus, if $m(x, y) < \epsilon$ and $m(x, z) > \epsilon$, there is an entourage that contains (x, y) but not (x, z) .

Thus, all pseudometrics induce pseudo-uniformities, but the converse is not true — there are pseudo-uniformities that are not induced by metrics. Two metrics m, m' on the same set of states induce the same uniformity iff the identity map is uniformly continuous in both directions.

Consider uniformities induced by pseudometrics on a fixed set of states \mathcal{S} .

Definition 8.2 \mathcal{M} is the class of pseudo-metrizable uniformities $\{\mathcal{U}_i\}$ on \mathcal{S} ordered as follows.

$$\mathcal{U}_1 \leq \mathcal{U}_2 \text{ if } \mathcal{U}_2 \subseteq \mathcal{U}_1$$

This order is closely related to the order on the lattice of pseudometrics.

Lemma 8.3 Let pseudometrics m_1, m_2 induce $\mathcal{U}_1, \mathcal{U}_2$ respectively, Let $\mathcal{U}_2 \subseteq \mathcal{U}_1$. Then pseudometric m defined as $m(s, t) = \max(m_2(s, t), m_1(s, t))$ also induces \mathcal{U}_1 .

Both the constructions of interest to us are “upto uniformities”.

Lemma 8.4 Let $(M, m), (M, m')$ be such that the uniformities induced by m, m' are the same. Then:

- $W(m), W(m')$ induce the same uniformity.
- $J(m), J(m')$ induce the same uniformity.

In the light of this theorem, we write $W(\mathcal{U})$ (resp. $J(\mathcal{U})$) for the uniformity generated by the constructions applied to a pseudo-metrizable uniformity \mathcal{U} . As a direct consequence of lemma 4.15.3, we have:

Corollary 8.5 *Let $\mathcal{U} \leq \mathcal{U}'$. Then $W(\mathcal{U}) \leq W(\mathcal{U}')$ and $J(\mathcal{S}_1) \leq J(\mathcal{S}_2)$.*

Combining the above results, we deduce the existence of a monotone function \mathcal{F}_k on the lattice of uniformities. This function is insensitive to the actual numerical value of the discount factor k .

Lemma 8.6 $(\forall 0 < k, k' < 1) \mathcal{F}_k = \mathcal{F}_{k'}$.

Since the lattice of uniformities is a complete lattice, \mathcal{F}_k has a complete lattice of fixed points. Thus, for any discount factor $0 < k < 1$, we get the same maximum fixed point in the lattice of uniformities. Furthermore:

Theorem 8.7 *The maximum fixpoint in (\mathcal{M}, \leq) is the uniformity induced by $m_{\mathcal{F}_k}$, the maximum fixpoint in (\mathcal{M}, \preceq) .*

The subtlety in the above proof is that in general greatest lower bounds in the lattice of uniformities does not coincide with those in the lattice of pseudometrics. Let $m_0 = \top$ and $m_{i+1} = \mathcal{F}_k(m_i)$ and let U_i be the uniformities induced by m_i . Then, the greatest lower bound of U_i is the uniformity induced by $\sup m_i$.

9. Examples

In this section, we discuss several examples of the use of approximate reasoning techniques. The general approach in this section is to identify natural quantitative observables, already explored in the literature, that are amenable to approximation — i.e. to calculate the observable at a state $\langle s, \tilde{c} \rangle$ upto ϵ , it suffices to calculate it a close-enough state $\langle s', \tilde{c}' \rangle$. This is clearly implied by continuity of the observable w.r.t. the metric $m_{\mathcal{F}_k}$.

The main technical tool that we use to establish continuity of observables is a continuous mapping theorem, e.g. see [26, 24] for an introductory exposition.

Theorem 9.1 (Continuous mapping theorem)

Let P_n be a sequence of probability distributions on X that converge to P . Let U be a function $X \rightarrow R$. Then $\int U dP_n$ converges to $\int U dP$.

9.1. Expected time to hit a proposition

Let \mathbf{p} be a proposition. We consider the expected time required to hit a p -state, i.e. a state where the proposition

\mathbf{p} is true. Define $\text{Hit}_p : \mathcal{D}^{m_F}[0, \infty) \rightarrow [0, \infty]$:

$$\text{Hit}_p(f) = \inf\{t \mid f(t) = \langle s, \tilde{c} \rangle, \mathbf{p} \text{ true at } s\}$$

Hit_p is a continuous function — if $J(m_F)(f, g) < \epsilon$, then $|\text{Hit}_p(f) - \text{Hit}_p(g)| < \epsilon$.

So, using the continuous mapping theorem, we deduce that if $\{\langle s_i, \tilde{c}_i \rangle\}$ converges to $\langle s, \tilde{c} \rangle$ then the sequence of expected times to hit a p -state from $\{\langle s_i, \tilde{c}_i \rangle\}$ converges to the expected time to hit a p -state from $\{\langle s, \tilde{c} \rangle\}$. In fact, since in this case, Hit_p is a 1-Lipschitz function, we can also deduce the rate of convergence using [26] — if $m_F(\langle s_i, \tilde{c}_i \rangle, \langle s, \tilde{c} \rangle) < \epsilon$, then the expected times to hit a p -state differ by at most 2ϵ .

9.2. Expected rewards

Let r_i be an assignment of rewards to states s_i such that if $r_i \neq r_j$, then states s_i, s_j differ in the truth-assignment of at least one proposition. This restriction can be viewed purely as a modelling constraint.

Define a function $R : \mathcal{G} \rightarrow [0, \infty)$ by:

$$R(\langle s_i, \tilde{c} \rangle) = r_i$$

Under the hypothesis that distinct rewards are distinguished propositionally, R defines a continuous function.

For any finitely-varying f , consider $\text{CumR}(f)$, a continuous function of t defined as follows:

$$\text{CumR}(f)(T) = \int_0^T R(f(t)) dt$$

By standard results — e.g. see [26] — CumR is a continuous function from $\mathcal{D}^{m_F}[0, \infty)$ to (C, unif) where C is the space of continuous functions from $[0, \infty) \rightarrow [0, \infty)$ with the uniform metric:

$$\text{unif}(f, g) = \sup_t |f(t) - g(t)|$$

Consider the following continuous functions from (C, unif) to $[0, \infty)$:

- For a fixed T , cumulative reward at time T .
- For a fixed T , average reward per unit time at T .
- The supremum of the times T at which cumulative reward is less than a fixed v , for some value v .
- The supremum of the times T at which the average reward is less than a fixed v , for some value v .

In each of these cases, by composing with CumR, we get a continuous function from $\mathcal{D}^{m_F}[0, \infty)$ to $[0, \infty)$. So, the continuous mapping theorem applies, and we deduce that if $\{\langle s_i, \tilde{\mathbf{c}}_i \rangle\}$ converges to $\langle s, \tilde{\mathbf{c}} \rangle$ then the sequence of expected values from $\{\langle s_i, \tilde{\mathbf{c}}_i \rangle\}$ converges to the expected value at $\langle s, \tilde{\mathbf{c}} \rangle$.

10. Functional characterization of uniformity

In this section, we provide an explicit construction of the maximum fixed point by considering a class of $[0, 1]$ valued functions. These functions ought to be viewed as the analogues of formulas in a real-valued modal logic.

10.1. Function expressions

Definition 10.1 Fix $0 < k < 1$. The syntax of function expressions is given by:

$$\begin{aligned} F_k^s &::= \mathbf{1} \mid \mathbf{p} \mid \min(F_k^s, F_k^s) \mid h \circ F_k^s \mid \int G_k^p \\ G_k^p &::= \mathcal{L}(F_k^s)(t) \mid \min(G_k^p, G_k^p) \mid h \circ G_k^p \end{aligned}$$

where \mathbf{p} ranges over atomic propositions, h is any Lipschitz operator on $[0, 1]$, $t \in [0, \infty)$.

Intuitively the F-function expressions are evaluated at generalized states, and the G-function expressions are evaluated on finitely-varying paths.

As preliminary intuition, $\mathbf{1}$ corresponds to the formula `true`, $\min(\cdot, \cdot)$ corresponds to conjunction², and $h \circ f$ encompasses both testing (via $h(x) = \max(x - q, 0)$) and negation (via $h(x) = 1 - x$). At a generalized state $\langle s, \tilde{\mathbf{c}} \rangle$, $\int G^p$ yields the (discounted) expectation of G^p wrt the distribution of $\text{Traces}(s, \tilde{\mathbf{c}})$. The intuition underlying $\mathcal{L}(F^s)(t)$ has been discussed in section 5 — at a finitely-varying function f , $\mathcal{L}(F^s)(t)$ yields the evaluation at time t of a time-smoothed variant of f .

We formalize these intuitions below. The interpretation of F-function expressions and G-function expressions yields maps whose range is the interval $[0, 1]$.

- The domain of F-function expressions is \mathcal{G} , the set of generalized states.
- The domain of G-function expressions is the set of finitely-varying functions with range \mathcal{G} .

² $\max(\cdot, \cdot)$ is definable as $1 - \min(1 - \cdot, 1 - \cdot)$ in both classes of function expressions.

Fix a GSMP. F-function expressions are evaluated as follows at a generalized state $\langle s, \tilde{\mathbf{c}} \rangle$:

$$\begin{aligned} \mathbf{p}(\langle s, \tilde{\mathbf{c}} \rangle) &= 1, \text{ iff } \mathbf{p} \text{ true at } s \\ \mathbf{1}(\langle s, \tilde{\mathbf{c}} \rangle) &= 1 \\ \min(F1_k^s, F2_k^s)(\langle s, \tilde{\mathbf{c}} \rangle) &= \min(F1_k^s(\langle s, \tilde{\mathbf{c}} \rangle), F2_k^s(\langle s, \tilde{\mathbf{c}} \rangle)) \\ h \circ F_k^s(\langle s, \tilde{\mathbf{c}} \rangle) &= h(F_k^s(\langle s, \tilde{\mathbf{c}} \rangle)) \\ \left(\int G_k^p \right)(\langle s, \tilde{\mathbf{c}} \rangle) &= k \times \int G_k^p(f) d\mu \end{aligned}$$

where μ is the distribution of $\text{Traces}(\langle s, \tilde{\mathbf{c}} \rangle)$.

G-function expressions are evaluated as follows at a finitely-varying function f :

$$\begin{aligned} \mathcal{L}(F_k^s)(f)(t) &= \sup_{t'} \{F_k^s(f(t')) - |t' - t|\} \\ \min(G1_k^p, G2_k^p)(f) &= \min(G1_k^p(f), G2_k^p(f)) \\ h \circ G_k^p(f) &= h(G_k^p(f)) \end{aligned}$$

Thus, for $f \in \text{Traces}(\langle s, \tilde{\mathbf{c}} \rangle)$, $\mathcal{L}(F_k^s)(f)(t)$ is the upper Lipschitz approximation to $F_k^s \circ f$ evaluated at t . We define a pseudometric d_k as follows.

Definition 10.2

$$d_k(\langle s, \tilde{\mathbf{c}} \rangle, \langle s', \tilde{\mathbf{c}}' \rangle) = \sup_{F_k^s} |F_k^s(\langle s, \tilde{\mathbf{c}} \rangle) - F_k^s(\langle s', \tilde{\mathbf{c}}' \rangle)|.$$

The uniformity induced by the d_k 's is independent of k and agrees with the maximal fixed point.

Theorem 10.3 *The uniformity induced by d_k coincides with the uniformity induced by $m_{\mathcal{F}_k}$, the maximum fixed point of \mathcal{F}_k .*

11. Conclusions

We have given a pseudo-metric analogue of bisimulation for GSMPs. We have shown that this really depends on the underlying uniformity and that quantities of interest are continuous in this metric. We have given a coinduction principle and a logical characterization reminiscent of previous work for weak bisimulation of a discrete time concurrent Markov chain.

The previous approaches to bisimulation work well for CTMCs, precisely because of the fact that the distribution is memoryless; at any given instant the expected duration in a state and the transition probabilities only depend on the current state of the system, and thus one can define a bisimulation on the state space. In contrast, the problem of describing bisimulation for real-time processes that have general distributions, rather than memoryless distributions, has been vexing. In the present work, we have shifted emphasis to the generalized states that incorporate time and not tried to define a bisimulation on the ordinary states. Because the generalized

states embody the quantitative temporal information we *have to work metrically*; an attempt to define bisimulation directly would have fallen afoul of the approximate nature of the timing information.

If we want to move to continuous state spaces and stochastic hybrid systems, the whole dynamical formalism has to be different: we have to use stochastic differential equations. That is a subject for future work.

References

- [1] R. Alur and D. Dill. A theory of timed automata. *Theoretical Computer Science*, 126:183–235, 1994.
- [2] R. Alur and T. Henzinger. Logics and models of real time: a survey. In J. W. de Bakker, C. Huizing, W. P. de Roever, and G. Rozenberg, editors, *REX workshop “Real time: Theory in Practice”*, volume 600 of *LNCS*, 1992.
- [3] E. J. Anderson and P. Nash. *Linear programming in infinite-dimensional spaces*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Ltd., 1987. Theory and applications, A Wiley-Interscience Publication.
- [4] C. Baier, J.-P. Katoen, and H. Hermanns. Approximate symbolic model checking of continuous-time markov chains. In *Proceedings of Concurrency Theory (CONCUR’99)*, volume 1664 of *Lecture Notes in Computer Science*. Springer-Verlag, 1999.
- [5] P. Billingsley. *Convergence of Probability Measures*. Wiley-Interscience, 1999.
- [6] M. Bravetti and R. Gorrieri. The theory of interactive generalized semi-markov processes. *Theoretical Computer Science*, 281(2):5–32, 2002.
- [7] F. Breugel and J. Worrell. An algorithm for quantitative verification of probabilistic systems. In *Proceedings of 12th International Conference on Concurrency Theory*, number 2154 in *Lecture Notes in Computer Science*, 2001.
- [8] F. Breugel and J. Worrell. Towards quantitative verification of probabilistic transition systems. In *Proceedings of the 28th International Colloquium on Automata, Languages and Programming (ICALP) 2001*, volume 2076 of *LNCS*, 2001.
- [9] R. Cleaveland, S. Smolka, and A. Zwarico. Testing preorders for probabilistic processes. In W. Kuich, editor, *Automata, Languages and Programming (ICALP 92)*, number 623 in *Lecture Notes in Computer Science*, pages 708–719. Springer-Verlag, 1992.
- [10] L. de Alfaro, T. Henzinger, and R. Majumdar. Discounting the future in systems theory. In *International Conference on Automata, Languages and Programming*, volume 2719 of *Lecture Notes in Computer Science*, pages 1022–1037. Springer, 2003.
- [11] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for partial labeled markov systems. In *International Conference on Concurrency Theory*, volume 1664 of *Lecture Notes in Computer Science*, 1999.
- [12] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. The metric analogue of weak bisimulation for probabilistic processes. In *Logic in Computer Science*, pages 413–422, 2002.
- [13] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labeled Markov systems. *Theoretical Computer Science*, 318:323–354, 2004.
- [14] R. Geroch. *Mathematical Physics*. University of Chicago Press, 1985.
- [15] H. Hansson. *Time and Probabilities in Formal Design of Distributed Systems*. Real-Time Safety Critical Systems Series. Elsevier, 1994.
- [16] B. R. Haverkort, L. Cloth, H. Hermanns, J.-P. Katoen, and C. Baier. Model checking performability properties. In *2002 International Conference on Dependable Systems and Networks (DSN 2002)*, 23-26 June 2002, Bethesda, MD, USA, *Proceedings*, pages 102–113. IEEE Computer Society, 2002.
- [17] H. Hermanns. *The Quest for Quantified Quality*, volume 2428 of *Lecture Notes in Computer Science*. Springer, 2002.
- [18] J. Hillston. *A Compositional Approach to Performance Modelling*. PhD thesis, University of Edinburgh, 1994. To be published as a Distinguished Dissertation by Cambridge University Press.
- [19] C.-C. Jou and S. A. Smolka. Equivalences, congruences, and complete axiomatizations for probabilistic processes. In J. Baeten and J. Klop, editors, *CONCUR 90 First International Conference on Concurrency Theory*, number 458 in *Lecture Notes in Computer Science*. Springer-Verlag, 1990.
- [20] K. G. Larsen and A. Skou. Bisimulation through probabilistic testing. *Information and Computation*, 94:1–28, 1991.
- [21] P. Lincoln, J. Mitchell, M. Mitchell, and A. Scedrov. A probabilistic poly-time framework for protocol analysis. In *ACM Computer and Communication Security (CCS-5)*, 1998.
- [22] A. Philippou, I. Lee, and O. Sokolsky. Weak bisimulation for probabilistic systems. In *Proceedings of CONCUR 2000*, *Lecture Notes in Computer Science*. Springer-Verlag, 2000.
- [23] R. Ségala and N. Lynch. Probabilistic simulations for probabilistic processes. *Nordic Journal of Computing*, 2(2):250–273, 1995.
- [24] G. S. Shedler. *Regeneration and networks of queues*. Springer-Verlag New York, Inc., 1987.
- [25] M. Vardi. Automatic verification of probabilistic concurrent finite-state systems. In *Foundations of Computer Science*, pages 327–338. IEEE Computer Society Press, 1985.
- [26] W. Whitt. *An Introduction to Stochastic-Process Limits and their Application to Queues*. Series: Springer Series in Operations Research. Springer-Verlag New York, Inc., 2002.
- [27] W. Whitt. Continuity of generalized semi markov processes. *Mathematics of Operation Research*, 5(4):494–501, 1980.