The Metric Analogue of Weak Bisimulation for Probabilistic Processes

Josée Desharnais*
Département d'Informatique
Université Laval
Québec, Canada, G1K 7P4
Josee.Desharnais@ift.ulaval.ca

Radha Jagadeesan†
Dept. of Computer Science
Loyola University-Lake Shore Campus
Chicago IL 60626, USA
radha@cs.luc.edu

Vineet Gupta
Stratify Inc.
501 Ellis Street
Mountain View CA 94043, USA
vineet@stratify.com

Prakash Panangaden‡
School of Computer Science
McGill University
Montreal, Quebec, H3A 2A7, Canada
prakash@cs.mcgill.ca

Abstract

We observe that equivalence is not a robust concept in the presence of numerical information - such as probabilities - in the model. We develop a metric analogue of weak bisimulation in the spirit of our earlier work on metric analogues for strong bisimulation. We give a fixed point characterization of the metric. This makes available coinductive reasoning principles and allows us to prove metric analogues of the usual algebraic laws for process combinators. We also show that quantitative properties of interest are continuous with respect to the metric, which says that if two processes are close in the metric then observable quantitative properties of interest are indeed close. As an important example of this we show that nearby processes have nearby channel capacities - a quantitative measure of their propensity to leak information.

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 inter-substituted for each other. Most investigations into probabilistic concurrent processes are also based on equivalences of one kind or another, e.g. [11, 15, 19, 13, 2, 16, 4, 22, 27, 21].

As is now recognized, this style of reasoning is too fragile in the sense of being too dependent on the exact numerical values of the probabilities[1]. This is particularly unfortunate for two reasons: firstly, the probabilities appearing in models really cannot be viewed as exact numbers; rather, they should be viewed as numbers with some error estimate. So, reasoning principles based on the exact value of numbers are of dubious practical value. Secondly, probability distributions over uncountably many states arise in even superficially discrete paradigms. For example, in the presence of recursion or iteration a binary choice inside a nonterminating computation determines an uncountable tree of possible outcomes with a probability distribution defined on it. Certainly in models of stochastic hybrid systems one has to deal with continuous state spaces. In such continuous-state systems with general probability distributions, an effective handle on the model is recovered by approximating with discrete probabilistic systems [8]. Clearly, these approximants do not match the continuous state model exactly. While our earlier approximation results show that working with finite-state systems is a good basis for dealing with general systems, it forces us to think about approximate reasoning principles.

Thus, we really want an "approximate" notion of

*Research supported by NSERC.
†Research supported in part by NSF .
‡Research supported in part by NSERC and MITACS.

[1] Indeed this is a criticism of the use of equivalences for any model that uses quantitative information in a serious way.

equality of processes. Furthermore, we want a compositional version of such an approximate equality, *i.e.*, we would like to move from the classical study of "exactly inter-substitutable" processes to the study of the more robust notion of "approximately inter-substitutable."

A natural pathway to approximate reasoning is to work with a relaxed notion of truth - for example using the interval $[0, 1]$ as the collection of truth values instead of $\{0, 1\}$. This is precisely Kozen's seminal idea on logics in the context of probability [17, 18]— moving from truth-valued boolean functions to real-valued (measurable) functions qua logical formulas. This demands that we take a more sophisticated view of the numerical probability values. This idea of the importance of numerical quantities also guides the search for a relaxation of the notion of equivalence of probabilistic processes. Jou and Smolka [16] note that the idea of saying that processes that are close should have probabilities that are close does not yield a transitive relation. This leads them to propose that the correct formulation of the "nearness" notion is via a metric.

Similar reasons motivate the study of Lincoln, Mitchell, Mitchell and Scedrov [20], our earlier studies of metrics for real-time systems [10] and our study of metrics for labelled Markov processes and strong bisimulation [7, 8] and the study of the fine structure of these metrics by van Breugel and Worrell [26, 25]. In contrast to these investigations, this paper will be carried out in the context of internal nondeterminism and weak bisimulation. The importance of weak bisimulation comes from the need for abstraction. In order to construct larger programs from smaller programs one works with the composition mechanisms of the language. When doing so it is necessary to hide internal actions and work with weak (rather than strong) bisimulation.

1.1 An Informal Summary of Results

We define a pseudometric on mixed nondeterministic and probabilistic processes such that zero distance corresponds to weak bisimilarity. Our main results are

- a fixed point characterization of this metric which permits coinductive proofs.
- a real-valued modal logic characterizing the metric.
- we show that several process combinators including parallel composition are non-expansive; this is an analogue of the congruence properties of weak bisimulation.
- a demonstration that nearness in the metric corresponds to nearness in quantitative properties of the processes - for example, we are able to generalize some common theorems about Markov chains and information theory to the case of concurrent Markov chains. This is of interest in its own right.

Working with nondeterminism and probability together (and with the metric analogue of weak bisimulation) forces us to confront some new technical obstacles. More specifically:

- Nondeterminism destroys the additivity property of the probability distributions. For each action we can now have multiple probabilistic distributions — letting the probability of a set of states be the maximum (or minimum) of these probabilities invalidates additivity.
- The algebraic laws of weak bisimulation disable the basic technique used to achieve contractive maps in the work on metric semantics, namely discounting future transitions and making them quantitatively less important than the current transitions. For example, $\tau.P$ is weakly bisimilar to P. Clearly, the "corresponding" transitions of P in both $P, \tau.P$ have to be weighted equally, even though they are delayed by a τ step in $\tau.P$. We overcome this hurdle by providing a novel explicit construction of the required (maximum) fixed point.

Criteria on Metrics for probabilistic processes A pseudometric[2] d on processes yields a real number distance for each pair of processes, *i.e.*, $d(P, Q) = d(Q, P)$ and $d(P, R) \leq d(P, Q) + d(Q, R)$. In our approach, we expect that the pseudometric d should satisfy:

- $d(P, Q) = 0 \Leftrightarrow P$ is weakly bisimilar to Q

This constraint can be viewed as a soundness criterion on the distance notion identified by the pseudometric.

The interesting cases of the distance function d are when it assigns a non-zero distance between two processes. We expect the pseudometric distance between processes to be roughly determined by the differences in the probabilities of similar transitions. For example, we expect $d(P_{\epsilon_1}, P_{\epsilon_2}) \leq |\epsilon_1 - \epsilon_2|$, where $P_\epsilon = a_{1-\epsilon}.Q$, is the process that does an a with probability $1 - \epsilon$ and then behaves like Q. Such criteria have good pedigree in the form of the properties

[2]Really we are working with pseudometrics on processes because $d(P, Q) = 0$ does not mean that $P = Q$.

of the Hutchinson metric on probability measures [14] on $[0,1]$ defined as $h(\mu,\nu) = \sup\{|\int f d\mu - f d\nu| \mid f$ a Lipschitz (non-expansive) function on [0,1]$\}$. This metric plays a fundamental role in fractals [9] and also in the work of van Breugel and Worrell [26, 25].

We define such a pseudometric d. We identify d as the maximum fixed point of a certain functional on pseudometrics. This approach exploits the associated good (co)universal properties and the associated coinduction proof rule permits us to recover much of the elegant proof techniques for (weak) bisimilarity. We follow up with an explicit construction of d — we consider a class $\mathcal{F}$ of $[0,1]$ valued functions on states, inspired by the logics for probabilistic transition systems [6]. $\mathcal{F}$ is to be viewed as a real-valued modal logic. These functions enable the definition of d as $d(P,Q) = \sup\{|f(s_P) - f(s_Q)| \mid f \in \mathcal{F}\}$ (s_P, s_Q are start states of P, Q respectively). The key technical tool in both these routes is a linear programming characterization of pseudometric distances inspired by the work of van Breugel and Worrell [25].

Process algebras provide the link to the desired compositional reasoning about approximate equality in such a pseudometric framework. If d is a pseudometric on processes, we would like *non-expansiveness*, expressed below as a desired compositional proof rule. For every n-ary process combinator $C[X_1,\ldots,X_n]$:

$$\frac{d(P_i,Q_i) \leq \epsilon_i,\ i = 1\ldots n}{d(C(P_1,\ldots,P_n), C(Q_1,\ldots,Q_n)) \leq \sum_i \epsilon_i}.$$

If we set $\epsilon_i = 0$ in the above equation, we get the condition that weak bisimulation is a congruence with respect to the operator $C[\cdot]$.

We show that parallel composition, restriction, CSP style internal choice $\oplus$ and guarded sum $a.P + b.Q$ - where a, b are not τ - in this language, are non-expansive in the sense of the above proof rule[3]. Our proofs are formally similar to the usual ones for pure labelled transition systems, showing that approximate reasoning (in our sense) is no harder than exact reasoning. Since weak bisimulation is *not* a congruence for CCS $+$ even in purely nondeterministic contexts, such a result does not hold for CCS $+$. Our extension of the theory to handle the CCS $+$ operator follows the spirit of the definition of observational congruence from weak bisimulation for labelled transition systems, and involves stricter matching of initial τ transitions. We show that the $+$ operator is non-expansive for this extension.

[3] Such a language is a probabilistic variant of languages for which weak bisimulation is known to be a congruence, for example languages in the ISOS format of Ulidowski [23, 24].

An application: Secure Substitution. The context for our investigations is exemplified by mobile code applications where programs (such as software for submitting income taxes) are downloaded as needed, executed on a trusted host (the home computer), require access to sensitive local data (such as financial information) and yet should not be permitted to "leak information". Thus, we are in the situation where a "mole" may have penetrated inside the system, and the analysis of the system has to account for the fact that this mole may be attempting to leak information to the outside world. Our interest is in *secure substitution* — when can one component (say C_2) can be substituted for another (say C_1) in the system without affecting the usual observations and the capability of the mole to leak information.

The definition of channel capacity in information theory offers a route to quantifying "information leakage". Channel capacity attempts to characterize the maximum amount of information that can be transmitted by a communication channel. In the context that we are considering, we are interested in measuring the information flow of the channel that is identified by the interface of the downloaded untrustworthy program. The (sequences of) labels on the internal interactions of this component with the system constitute the input symbols of the channel. The "mole" is viewed as attempting to leak information about the sequences of these input symbols by interacting with the outside world, the (sequences of) labels on these interactions constitute the output symbols. The outside observer is attempting to deduce input traces while only viewing the output traces.

The first technical obstacle that we need to overcome is that in the presence of nondeterminism, we are working with sets of distributions rather than single distributions. We model worst case assumptions by permitting the mole to control the (nondeterministic) scheduler, by permitting the mole to serve as (a possibly probabilistic) oracle to guide the scheduler[4]. Our definition of channel capacity captures these worst case assumptions.

Given this modelling assumption, we describe the following results. Using basic results from information theory and exploiting the non-expansive properties, we show that there is a constant k only dependent on C_1 such that: $d(C_1, C_2) < \epsilon$ implies that the difference in

[4] Our modelling differs from the study of Lincoln, Mitchell, Mitchell and Scedrov [20] in this aspect. They use a probabilistic process algebra to explore the power of arbitrary probabilistic polynomial time processes in the context of security protocols. In their model, the system is probabilistically determinate and the adversary is "outside" the system and thus has no useful control over the scheduler of the system. Rather, it derives its power from the ability to intercept and alter messages.

the channel capacities of C_1, C_2 is at most $k\epsilon$. Thus, the channel capacity is a continuous function of the pseudometric. This analysis provides an à posteriori justification for the criteria on pseudometrics discussed earlier. This result assumes particular force in conjunction with the non-expansiveness of the basic process combinators and gives us the basic rudiments of a methodology to determine if a component can be securely substituted for another.

An application: Quantitative reasoning Very often one wants not just logical properties but quantitative properties of systems. For example, one more likely wants to know the average size of a queue rather than, say, the fact that it is nonempty with probability one. Similarly one often wants to know large deviation probabilities away from the mean.

Consider a queueing system with multiple servers and a stream of customers arriving with a given distribution. There are two basic strategies: one can have all the customers in a single queue with each server taking the next customer from the common queue (as is typical in banks) or one could have a queue for each server with the customer committing to a queue upon entry to the system (as in supermarkets). Given distributions for arrival times and service times we would like to know if these systems are close in their quantitative properties such as expected waiting times, average throughput. Clearly the two systems are quite different in terms of their exact logical properties and one would get no sense of how different they are if we were to compare them on that basis. In terms of modelling with a probabilistic process algebra, we cannot really make comparisons using a metric analogue of strong bisimulation because the transitions are very different right away. However, if we hide the internal details of how processes are allocated to different queues and discount these internal transitions we would have the basis for a quantitative comparison.

It might appear, at first sight, that abstracting from the internal transitions would destroy the temporal information. However, time really passes while waiting in the queues and while interacting with the servers. One can set up a timer process to keep track of the times. What one is really abstracting over are the internal transitions that govern the interactions with the queue(s). It is not hard to show that such average quantitative properties are continuous which means that if two systems are metrically close their quantitative properties will be close as well. Thus the metric distance gives a good measure of how closely the quantitative properties of two systems match.

2. Definitions

We begin with a review of the underlying framework — our definitions are adapted from [21]. We work in the context of the "alternating model" for labelled concurrent Markov chains [12], labelled transition systems with non-determinism and probability.

Definition 2.1 *A labelled concurrent Markov chain (henceforth LCMC), is a tuple $\mathcal{K} = (K, \mathtt{Act}, \longrightarrow, k_0)$, where*
(1) $K = K_p \cup K_n$, a countable set, is partitioned into the probabilistic states, K_p, and the nondeterministic states K_n. The start state is $k_0 \in K$.
(2) $\mathtt{Act}$ is a finite set of action symbols that contains a special action τ.
(3) The transition relation $\longrightarrow = \longrightarrow_p \cup \longrightarrow_n$ is partitioned into probabilistic and nondeterministic transitions. $\longrightarrow_n \subseteq K_n \times \mathtt{Act} \times K_p$ is image-finite, i.e. *for each $s \in K_n$ and $a \in \mathtt{Act}$, the set $\{s' \in K_p \mid s \stackrel{a}{\to} s'\}$ is finite. $\longrightarrow_p \subseteq K_p \times (0,1] \times K_n$ satisfies that for each $s \in K_p$ $\sum_{(s,\pi,t)\in\longrightarrow_p} \pi \leq 1$.*

Thus, transitions from nondeterministic states are finitely branching[5] and labelled. Transitions from probabilistic states are un-labelled and are associated with numbers that are interpreted as probabilities. *In this paper, we will work with finite state systems,* the domain of definition of the weak bisimulation definition of Lee, Philippou and Sokolsky [21].

Every probabilistic state s induces a distribution P on K given by $P(t) = \sum_{(s,\pi,t)\in\longrightarrow_p} \pi$ for every $t \in K$. We sometimes write $s \to_p P$ to emphasize this distribution.

Example 2.2 (Constructions on LCMCs)
Given LCMCs $\mathcal{K}_i$, the LCMC $\sum_i a_i.\mathcal{K}_i$, where $a_i, i = 1, \ldots n$ is a finite collection of labels, is constructed by adding an a_i-transition to $\mathcal{K}_i$. Assume that the $\mathcal{K}_i$'s have pairwise disjoint state spaces and let $\longrightarrow_i$ be their transition relations and s_i their start states, which are assumed to be probabilistic. Let $s \notin \cup K_i$. Then, $\sum_i a_i.\mathcal{K}_i = (\cup K_i \cup \{s\}, \mathtt{Act}, \bigcup_i \longrightarrow_i \cup \{s \stackrel{a_i}{\to} s_i \mid i = 1, \ldots n\}, s)$.
Given LCMCs $\mathcal{K}_1, \mathcal{K}_2$, the LCMC $\mathcal{K}_1 + \mathcal{K}_2$ is constructed as follows. Formally, assume that $\mathcal{K}_1$ and $\mathcal{K}_2$ have disjoint sets of states except for the start state k which is nondeterministic and let $\longrightarrow_i$ be the transition relations of $\mathcal{K}_i$, $i = 1, 2$. Then, $\mathcal{K}_1 + \mathcal{K}_2 = (K_1 \cup K_2, \mathtt{Act}, \longrightarrow_1 \cup \longrightarrow_2, k)$.

[5] Since we have a finite action set, "image finite" and "finitely branching" are the same thing.

We use some notation for sequences (of states or transitions). We use ε for the empty sequence and $\cdot$ for concatenation. Every sequence, say σ, of transitions has an associated probability $\mathtt{prob}(\sigma)$, obtained by multiplying the probabilities occurring on the path. Thus, we attribute probability 1 to a nondeterministic transition in a path, and multiply in the probability of a probabilistic transition. Similarly, every sequence σ of transitions has an associated weak sequence of labels $\mathtt{Weak}(\sigma) \in (\mathtt{Act} - \{\tau\})^*$, obtained by removing the τ's. Thus probabilistic transitions and nondeterministic transitions with label τ do not contribute to the weak label.

We define *computations* of an LCMC as transition trees obtained by unfolding the LCMC from the root, resolving the nondeterministic choices (*i.e.* each nondeterministic state has at most one transition coming out of it) and taking all probabilistic choices at a probabilistic state. A computation can thus be viewed as a purely (sub)probabilistic labelled Markov chain. We elide standard definitions of trees.

Definition 2.3 *Let $\mathcal{K}$ be a LCMC, $a \in \mathtt{Act}$. An a-computation from $s \in K$ is a computation such that every path from the root has weak label a or ε.*

We allow a ε label in order to allow paths that could be extended to have an a-transition in an extended a-computation.

Each computation induces a distribution on its leaf states in the standard way — the probability of a leaf node is the probability of the (unique) path going to it. Here we insist that the paths contributing to the distribution do have an a-label.

Definition 2.4 *Let $\mathcal{K}$ be a LCMC, $s \in K$, and let Q be a distribution on states.*
We write $s \stackrel{a}{\Rightarrow} Q$, if there is an a-computation such that for all $s_i \in K$, $Q(s_i) \leq \sum_\sigma \mathtt{prob}(\sigma)$ where the summation is taken over paths σ with weak label a that start in s and end in the leaf s_i.

We extend this notation to linear combinations of Q's. Let $s_i \stackrel{a}{\Rightarrow} Q_i$, and let $\sum_i \lambda_i = 1$. Then, we write: $\sum_i \lambda_i \times (s_i \stackrel{a}{\Rightarrow} Q_i)$. In the special case where all $s_i = s$, we write $s \stackrel{a}{\Rightarrow} \sum_i \lambda_i \times Q_i$.

$s \stackrel{a}{\Rightarrow} [\lambda \times P + (1-\lambda) \times Q]$ is reminiscent of the randomized schedulers [22].

We define the "probability" from a state s to a subset of states via a path with weak label a by taking the supremum over all possible computations.

Definition 2.5 *Let $\mathcal{K}$ be a LCMC, $s \in K, E \subseteq K$. Then, the probability of going from s to E via a, denoted by $P(s,a,E)$, is defined as:*

$$P(s,a,E) = \sup\{\sum_{s \in E} Q(s) \mid s \stackrel{a}{\Rightarrow} Q\}.$$

The supremum in this definition is the source of the subtlety of weak bisimulation — $P(s,a,.)$ does not satisfy additivity. We are now ready to define weak bisimulation. Our presentation of this definition is different from [21] — in the full paper we show that it does indeed coincide with [21] for finite state systems. We consider equivalence relations on the set of states. Given an equivalence relation R, we say a set E is R-closed if $E = \{s \mid \exists t \in E \text{ such that } tRs\}$.

Definition 2.6 *An equivalence relation R on K is a weak bisimulation if for all $s, t \in K$ such that $s\ R\ t$ and all R-closed $E \subseteq K$, we have:*

$$(\forall a \in \mathtt{Act})\ [P(s,a,E) = P(t,a,E)].$$

There is a maximum weak bisimulation, denoted by $\approx$.

The equational laws supported by this definition extend the usual ones for nondeterministic labelled transition systems or purely probabilistic transition systems. Indeed, the usual relations that witness the bisimulation are carried over essentially unchanged, for example, $\tau.\mathcal{K} \approx \mathcal{K}$, and unfolding a LCMC yields a weakly bisimilar system. See [3] for a full axiomatization of equational laws for finite processes.

3. The Pseudometric as a maximum fixed point

Our first presentation of the pseudometric is as a maximum fixed point of a certain functional.

We fix an $LCMC$ and consider pseudometrics on its set of states.

Definition 3.1 *$\mathcal{M}$ is the class of 1-bounded pseudometrics on states with the ordering*

$$m_1 \preceq m_2 \text{ if } (\forall s,t)\ [m_1(s,t) \geq m_2(s,t)]$$

Lemma 3.2 *$(\mathcal{M}, \preceq)$ is a complete lattice.*

Proof. The least element is given by: $\bot(s,t) = 0$ if $s = t$, 1 otherwise. The top element is given by $(\forall s,t) \top(s,t) = 0$. Greatest lower bounds are given by: $(\sqcap\{m_i\})(s,t) = \sup_i m_i(s,t)$, which can be shown to satisfy the triangle inequality. ∎

$m \in \mathcal{M}$ is extended to distributions on sets of states as follows. The definition is based on the Hutchinson metric on probability measures — we have merely simplified the definition for our context of discrete finite state distributions.

Definition 3.3 *Let $m \in \mathcal{M}$. Let P, Q be distributions on states. Then $m(P,Q)$ is given by the solution to the following linear program:*

$$\max \sum_i (P(s_i) - Q(s_i))(a_i - a_K)$$
$$\textit{subject to}: \quad \forall i. 0 \leq a_i, a_K \leq 1$$
$$\forall i, j.\ a_i - a_j \leq m(s_i, s_j).$$

We need the constraints $a_i \leq 1$ if the distributions do not have equal total probability, without them the maximum is unbounded. Equivalently, $m(P,Q)$ is given by the solution to the following dual linear program:

$$\min \sum_{i,j} l_{ij} m(s_i, s_j) + \sum_i x_i + \sum_j y_j$$
$$\text{subject to}: \quad \forall i. \textstyle\sum_j l_{ij} + x_i = P(s_i)$$
$$\forall j. \textstyle\sum_i l_{ij} + y_j = Q(s_j)$$
$$\forall i, j.\ l_{ij}, x_i, y_j \geq 0.$$

This extension of pseudometrics of $\mathcal{M}$ to distributions satisfies symmetry and the triangle inequality, *i.e.* $(\forall P, Q, R)\ m(P,Q) \leq m(P,R) + m(Q,R)$. It is also consistent with the ordering on pseudometrics, *i.e.* $m \preceq m' \Rightarrow (\forall P, Q)\, m(P,Q) \geq m'(P,Q)$.

The dual linear program is a key tool to move from distributions to states in the following sense. Given close-by distributions P, Q, the dual linear program permits us to construct a matching of states (that may include "splitting" of the probabilities assigned by P, Q), in such a way that exactly the distance between P, Q can be recovered.

Lemma 3.4 *Given distributions P, Q, there exists distributions P_i, Q_i such that*

- *P_i are point distributions that are non-zero at only one state.*
- *$P = \sum P_i$, $Q = \sum Q_i$*
- *$m(P,Q) = \sum m(P_i, Q_i)$.*

We now define a functional F on $\mathcal{M}$ that closely resembles the usual functional for weak bisimulation.

Definition 3.5 *Define F, a functional on $\mathcal{M}$ as follows. $F(m)(s,t) < \epsilon$ if:*

- $(\forall s \stackrel{a}{\Rightarrow} P)\ (\exists t \stackrel{a}{\Rightarrow} Q)\ [m(P,Q) < \epsilon]$.
- $(\forall t \stackrel{a}{\Rightarrow} Q)\ (\exists s \stackrel{a}{\Rightarrow} P)\ [m(P,Q) < \epsilon]$.

$F(m)(s,t)$ is thus given by:

$$\begin{array}{ll}\max(& \max_{a \in \mathtt{Act}} \sup_{s \stackrel{a}{\Rightarrow} P} \inf_{t \stackrel{a}{\Rightarrow} Q} m(P,Q), \\ & \max_{a \in \mathtt{Act}} \sup_{t \stackrel{a}{\Rightarrow} Q} \inf_{s \stackrel{a}{\Rightarrow} P} m(P,Q))\end{array}$$

F is monotone on $\mathcal{M}$. Using Tarski's fixed point theorem, F has a maximum fixed point. Using the image finiteness of LCMC, we can show that the closure ordinal of F is ω. The proof proceeds standardly, by showing that the maximum fixed point m is given by $m = \sqcup_i m_i$, where $m_0 = \top$ and $m_{i+1} = F(m_i)$.

The maximal fixed point of F is sound with respect to bisimulation. The forward implication of the proof uses the pseudometric m' defined as: $m'(s,t) = 0$ iff s and t are bisimilar and 1 otherwise. m' satisfies $m' \preceq F(m')$. The converse proceeds by showing that the equivalence relation R induced by 0 distance is a bisimulation.

Lemma 3.6 $s \approx t \Leftrightarrow m(s,t) = 0$*, where m is the maximum fixed point of F.*

Finally, we show that it suffices to consider one-step transitions for the hypothesis in the definition of the functional F. F' demands only matching of "one-step" transitions. However, by using lemma 3.4 to move from distributions to states, we can show that F' enforces the matching required by F. Let δ_s means the Dirac measure at s.

Lemma 3.7 *Define $F' : \mathcal{M} \to \mathcal{M}$ as follows: $F'(m)(s,t) < \epsilon$ if:*

- $(\forall s \stackrel{a}{\to} s')\ (\exists t \stackrel{a}{\Rightarrow} Q)\ [m(\delta_{s'}, Q) < \epsilon]$.
  $(\forall s \to_p P)\ (\exists t \stackrel{\tau}{\Rightarrow} Q)\ [m(P,Q) < \epsilon]$.
- $(\forall t \stackrel{a}{\to} t')\ (\exists s \stackrel{a}{\Rightarrow} P)\ [m(P, \delta_{t'}) < \epsilon]$.
  $(\forall t \to_p Q)\ (\exists s \stackrel{\tau}{\Rightarrow} P)\ [m(P,Q) < \epsilon]$.

Then, the maximum fixed points of F, F' coincide.

Explicit construction of pseudometric. We now 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.

Definition 3.8 $\mathcal{F}$ is a set of function expressions whose syntax is given by:

$$f ::= \mathbf{1} \mid \max(f,f) \mid h \circ f \mid a.f$$

where a is a label, possibly τ and h is any non-expansive operator on [0,1] ($|h(x) - h(y)| \leq |x - y|$). Given an LCMC $(K, \texttt{Act}, \longrightarrow, k_0)$, these function expressions are evaluated on K as follows:

$$\begin{array}{rcl} \mathbf{1}(s) & = & 1 \\ \max(f_1, f_2)(s) & = & \max(f_1(s), f_2(s)) \\ h \circ f(s) & = & h(f(s)) \end{array}$$

If $a \neq \tau$,

$$a.f(s) = \begin{cases} \sum_i P(s_i) a.f(s_i), \text{if } s \to_p P \\ \max(\ \{\tau.f(s') | s \xrightarrow{a}_n s'\} \cup \\ \qquad \{a.f(s') | s \xrightarrow{\tau}_n s'\}) \\ \qquad \text{if } s \in K_n. \end{cases}$$

Finally,

$$\tau.f(s) = \begin{cases} \max(f(s), \{\tau.f(s') | s \xrightarrow{\tau}_n s'\}), s \in K_n \\ \max(f(s), \sum_i P(s_i)\tau.f(s_i)), s \to_p P. \end{cases}$$

These functions are inspired by a simple modal logic. $\mathbf{1}$ corresponds to the formula `true`, $\max(f_1, f_2)$ corresponds to disjunction, and $a.f$ corresponds to the next modality operator. $h \circ f$ encompasses both testing and negation — $h(x) = 1 - x$ corresponds to negation, whereas $h(x) = x - q$ if $x > q$, 0 otherwise allows testing the probability values. We define a pseudometric d as follows.

Definition 3.9 $d(s,t) = \sup_{f \in \mathcal{F}} |f(s) - f(t)|$.

d coincides with m.

Theorem 3.10 *d is the maximum fixed point of F.*

4. Process algebra

The usual proofs to show that weak bisimulation is a congruence for prefixing, internal/probabilistic choice and restriction essentially work in the metric setting.

Lemma 4.1 *Prefixing, internal choice, probabilistic choice and restriction are non-expansive.*

Proof. (Sketch) In this abstract, we consider only prefixing. The construction have already been described in example 2.2. Let m be the pseudometric witnessing $d(s_1, s_2) < \epsilon$. We will construct the required witness m' such that $m' \preceq F'(m')$ and $m'(u_1, u_2) < \epsilon$, where u_i are the start states of $a.s_i$. Define $m'(u_1, u_2) = m(s,t)$. For all other states $x \neq u_1, u_2$, $m'(u_1, x) = m'(u_2, x) = 1$. Finally, $m'(x,y) = m(x,y)$ for all states $x, y \notin \{u_1, u_2\}$. ∎

Parallel composition. For any LCMC, we can construct a bisimilar "saturated" transition system, that has the property that every nondeterministic state is bisimilar to some probabilistic state and vice versa. Given $\mathcal{K} = (K, \texttt{Act}, \longrightarrow, k_0)$, construct $\mathcal{U} = (U, \texttt{Act}, \longrightarrow, u_0)$ as follows: replace each transition $s \xrightarrow{a} t$ from a nondeterministic state s by a sequence $s \xrightarrow{a} s_1$, $s_1 \xrightarrow{1} s_2$, $s_2 \xrightarrow{\tau} t$, where s_1, s_2 are new states not in K. Then, for every nondeterministic state s add the following self-loop: $s \xrightarrow{\tau} s^p$, $s^p \xrightarrow{1} s$, where s^p is again a new state.

We will now define the parallel composition of LCMCs. Assume that both LCMCs are saturated and that their start states are nondeterministic. When restricted to LTSs, our definition agrees upto weak bisimilarity. On saturated LCMCs, our definition agrees upto strong bisimulation with the definition of Hansson and Johnson [11] as modified in the thesis of Hansson [12].

Given saturated LCMCs $\mathcal{K}, \mathcal{L}$ whose start states are nondeterministic, we define $\mathcal{M} = \mathcal{K} || \mathcal{L}$ as follows. For any nondeterministic state s (in either of $\mathcal{K}, \mathcal{L}$), we use s^p to refer to the corresponding bisimilar probabilistic state to which s has a τ transition. The nondeterministic (resp. probabilistic) states of $\mathcal{M}$ are given by $K_n \times L_n$ (resp. $K_p \times L_p$). The transition relation is:

$$\begin{array}{ll} (s,t) \xrightarrow{a} (s', t^p), (t,s) \xrightarrow{a} (t^p, s') & \text{if } s \xrightarrow{a} s' \\ (s,t) \xrightarrow{\tau} (s', t') & \text{if } s \xrightarrow{a} s', t \xrightarrow{\bar{a}} t' \\ (s,t) \xrightarrow{p \times q} (s', t') & \text{if } s \xrightarrow{p} s', t \xrightarrow{q} t' \end{array}$$

This definition preserves saturation and strict alternation of nondeterministic and probabilistic transitions.

Lemma 4.2 *Parallel composition is non-expansive.*

Proof. (Sketch) Consider $\mathcal{K}_1 || \mathcal{L}$ and $\mathcal{K}_2 || \mathcal{L}$. Let m be the pseudometric witnessing $d(s_1, s_2) < \epsilon$, where $s_i \in \mathcal{K}_i$. Define $m'((s_1,t),(s_2,t)) = m(s_1,s_2)$, where $s_i \in \mathcal{K}_i$, both s_1, s_2 nondeterministic or probabilistic, and $m'((s_1,t),(s_2,t)) = 1$ otherwise. Then m' satisfies $m'(u_1,u_2) < \epsilon$ where u_i is the start state of $\mathcal{K}_i || \mathcal{L}$, and $m' \preceq F'(m')$. ∎

Handling CCS $+$. We now sketch the extension of the theory to handle the CCS $+$ operator. Weak bisimulation is not a congruence for CCS $+$ operator even in LTSs, a situation remedied by a stricter matching of initial τ transitions.

We follow this standard trick in the metric context. We write $s \stackrel{\tau^+}{\Rightarrow} P$ for a computation P from s such that every path from s has at least one τ transition. The d^c

pseudometric agrees with d wrt matching of all non-τ transitions. For τ-transitions, we demand a match by a transition of the form $s \stackrel{\tau^+}{\Rightarrow} P$.

Definition 4.3

$$d^c(s,t) = \max \begin{cases} d(s,t), \\ \max(\sup_{P\in A} \inf_{Q\in B} d(P,Q), \\ \qquad \sup_{Q\in B} \inf_{P\in A} d(P,Q)) \end{cases}$$

where $A = \{P \mid s \stackrel{\tau^+}{\Rightarrow} P\}, B = \{Q \mid t \stackrel{\tau^+}{\Rightarrow} Q\}$.

Zero distance in d^c agrees with the largest congruence contained in weak bisimulation [3]. Thus, if $d^c(s,t) = 0$, for all one-step τ transitions $s \stackrel{\tau}{\to} P$ (nondeterministic or probabilistic) , there exists $t \stackrel{\tau^+}{\Rightarrow} Q$ such that $d(P,Q) = 0$.

Lemma 4.4 $+$ *is non-expansive wrt* d^c.

For the real-valued modal logic, the addition of a top-level τ^+ test is exactly what is needed to enable the real-valued modal logic to capture d^c.

$$\tau^+.f(s) = \begin{cases} \max(\{\tau.f(s') | s \stackrel{\tau}{\to}_n s'\}) & \text{if } s \in K_n \\ \max(\sum_i P(s_i)\tau.f(s_i)) & \text{if } s \to_p P. \end{cases}$$

Lemma 4.5

$$d^c(s,t) = \max \begin{cases} \sup_{f\in\mathcal{F}} |\tau^+.f(s) - \tau^+.f(t)| \\ \sup_{f\in\mathcal{F}} |f(s) - f(t)| \end{cases}$$

5. Secure substitution

The context for our investigations is exemplified by mobile code applications where programs (such as tax software) are downloaded as needed, executed on a trusted host (the home computer), require access to sensitive local data (such as financial information) and yet should not be permitted to “leak information”. Thus, we are in the situation where a “mole” has been permitted inside the system and we are interested in measuring the information that the mole can leak to the outside world. We use the definition and basic results about channel capacity from information theory — see [5] for details.

We first formalize the interface of this channel. Fix an LCMC, $\mathcal{K} = (K, \texttt{Act}, \longrightarrow, k_0)$. Fix $O \subset \texttt{Act} - \{\tau\}$, a subset of labels not including τ that is intended to model the “output labels” visible to the external observer. The remaining labels, $I = (\texttt{Act} - \{\tau\}) \setminus O$ is the set of “input labels”. The (sequences of) labels from I constitute the input symbols of the channel. The “mole” is viewed as attempting to leak information using these input symbols by influencing the (sequences of) O labels. These sequences of output labels constitute the output symbols. Thus, the outside observer is attempting to deduce input traces (elements of $I^\star$) while only viewing the output traces (elements of $O^\star$).

Next, we identify the probabilistic transducers from input symbols to output symbols associated with this channel. Inevitably, in the presence of nondeterminism, we cannot identify a single transducer, rather we are forced to accommodate a set of probabilistically determinate transducers. The following definition of O-determinate subtransition systems captures the conditions on the elements of this set. The conditions reflect two aims. First, we want to ensure liveness conditions that prevent the adversary from blocking the system from progress, if it is possible. Secondly, we want to ensure that the nondeterminism is resolved enough to get a (probability) distribution on output symbols once the input symbols are fixed.

First some notation. For purely notational convenience, assume that the LCMC is completely unrolled into a tree. We say that two sequences of transitions σ, σ' are *consistent*, if the choices at nondeterministic states that occur in both σ, σ' are consistent. For a state s and a subset of transitions T , define $Enabled^T(s) = \{a | a \neq \tau$, there is a path with weak label a from s in $T\}$. We will consider subsets T of transitions that satisfy:

1. Liveness: For all s, $|Enabled^T(s) \cap \texttt{Act}| = 0 \Rightarrow |Enabled^{\longrightarrow}(s) \cap \texttt{Act}| = 0$.
2. I-liveness: For all s such that $|Enabled^T(s) \cap O| = 0$, $Enabled^T(s) \cap I = Enabled^{\longrightarrow}(s) \cap I$.
3. O-determinacy: Let s be such that $|Enabled^T(s) \cap O| \geq 1$. Let $a, b \in \texttt{Act} \setminus \{\tau\}$. Let $\sigma_1 = s \stackrel{\tau^*}{\to} s_1' \stackrel{a}{\to} s_1$ and $\sigma_2 = s \stackrel{\tau^*}{\to} s_2' \stackrel{b}{\to} s_2$ be sequences of transitions in T. Then, σ_1, σ_2 are consistent.
4. I-determinacy: Let $a \in I$, s a state. Let $\sigma_1 = s \stackrel{\tau^*}{\to} s_1' \stackrel{a}{\to} s_1$ and $\sigma_2 = s \stackrel{\tau^*}{\to} s_2' \stackrel{a}{\to} s_2$ be sequences of transitions in T. Then, σ_1, σ_2 are consistent.

The first condition ensures that T cannot reject all labels. The second condition states that the only reason for T to reject an input symbol (that might otherwise be accepted) at a state is a nondeterministic choice leading to an output symbol. The third condition ensures that the choices at states that can perform a weak O-labelled transition are purely probabilistic. The last condition ensures that at any state, there is only one computation for a given I-label.

The transition systems satisfying these conditions induce probabilistic transformers from sequences (say of length m) of I-labels to sequences of O-labels (say of length n). Let $\sigma \upharpoonright A, A \subseteq \texttt{Act}$ denote the subsequence of A-actions in $\texttt{Weak}(\sigma)$.

Definition 5.1 *Let $T \subseteq \longrightarrow$ satisfy the above conditions. Let $x \in I^m, y \in O^n$. Define:*

$$\begin{aligned} p_T^{m,n}(y|x) &= \sum\{prob(\sigma) \mid \sigma = s \rightarrow s_1 \rightarrow \ldots \xrightarrow{a} t, \\ &\qquad a \neq \tau, \sigma \upharpoonright O = y, \sigma \upharpoonright I = x\} \\ p_T^{m,n}(\delta|x) &= 1 - \sum_{y \in O^n} p_T^{m,n}(y|x) \end{aligned}$$

δ is a special symbol to indicate absence of output from O^n. $p_T^{m,n}(.|x)$ is a probability distribution on $O^n \cup \{\delta\}$.

The channel capacity of a single probabilistic transformer such as the one above is defined in a standard fashion — see [5] for detailed intuitions.

Definition 5.2 *Given a joint probability distribution $f(x,y)$, define mutual information, written $I(f)$ as:*

$$I(f) = -\sum_{x \in X} \sum_{y \in Y} f(x,y) \log[\frac{f(x,y)}{f(x) \times f(y)}].$$

The channel capacity of $p(y|x)$, a probabilistic transducer from inputs x to outputs y, written $Ch(p)$, is $\max_r I(g)$ where $r(x)$ is a probability distribution on inputs x and $g(x,y) = r(x) \times p(y|x)$.

The general theorems of information theory guarantee that the channel capacity calculated as per the above definition can be achieved operationally by repeated use of the transducer p. This definition can of course be used on $p_T^{m,n}((.)|x)$ defined above since it is a pure probabilistic transducer.

Recall that we model worst case assumptions about the mole by allowing the mole to control the (nondeterministic) scheduler and by permitting it to serve as oracle to guide the scheduler.

Definition 5.3 *The O-channel capacity of a state s, written $Ch(s)$ is defined as the supremum over all m, n, T of the channel capacity of $p_T^{m,n}((.)|x)$.*

As per this definition, the mole can choose a nondeterministic scheduler to derive a purely probabilistic computation that gets arbitrarily close to the value of the O-channel capacity of a state s. The mole does not gain anything by using probabilities to combine several such purely probabilistic computations because of the convexity of mutual information:

Theorem 5.4 ([5]) *Let $p_1(y|x), p_2(y|x)$ be probabilistic transducers from inputs x to outputs y. Let $r(x)$ be any distribution on inputs. Let $g_i(x,y) = r(x) \times p_i(y|x)$, for $i = 1, 2$. Let $0 \leq \lambda \leq 1$. Then $I(\lambda g_1 + (1-\lambda) g_2)) \leq \lambda I(g_1) + (1-\lambda) I(g_2)$*

Thus, mixing probabilistic transducers diminishes the channel capacity, validating our definition 5.3 as the correct modelling of the worst case assumptions.

Basic information theory also provides bounds on changes in channel capacity as a function of changes in probability for purely probabilistic transducers.

Lemma 5.5 *Let $p(y|x), p'(y|x)$ be probabilistic transducers from inputs to outputs such that $\max_{x,y} |p(y|x) - p'(y|x)| < \epsilon$. Then: $|Ch(p) - Ch(p')| < k\epsilon$, for some constant k that depends only on p.*

The coinduction infrastructure that we have described earlier permits us to match nondeterministic choices between nearby processes. Thus, if s, t are close, every probabilistically determinate computation from s that contributes to the channel-capacity of s can be matched by one from t. This permits us to lift the bound described above for single probabilistic transducers to sets of probabilistic transducers and reach our goal of showing that nearby states have almost the same channel capacities.

Theorem 5.6 *Let $d(s,t) < \epsilon$. Then $|Ch(s) - Ch(t)| < k\epsilon$, for some constant k that depends only on s. Thus, $Ch(\cdot)$ is a continuous function w.r.t. pseudometric d.*

6. Conclusions

We have described a fixed-point approach to a metric analogue for weak bisimulation. This is the first time to our knowledge that the "ϵ-analogue" of the usual process algebra equations for weak (or strong for that matter) bisimulation have been developed. The fixed point approach was crucial to this development since it made coinductive techniques possible. We were able to make use of the beautiful duality principle from linear programming.

We have also explored the quantitative meaning of the metric. In particular we have shown that the analogue of the information theory analysis of Markov chains can be extended to this setting (*i.e.* to concurrent Markov chains). This involved an extension of the usual concepts like channel capacity.

It may appear that our use of linear programming forces us into the setting of finite state systems. However, the notion of duality works well for infinite-dimensional spaces [1] and we are exploring this extension.

References

[1] E. J. Anderson and P. Nash. *Linear Programming in Infinite-dimensional Spaces*. Discrete Mathematics and Computation. Wiley-Interscience, 1987.

[2] J. Baeten, J. Bergstra, and S. Smolka. Axiomatizing probabilistic processes: ACP with generative probabilities. *Information and Computation*, 121(2):234–255, 1995.

[3] E. Bandini and R. Segala. Axiomatizations for probabilistic bisimulation. In *Proceedings of the 28th International Colloquium on Automata, Languages and Programming*, number 2076 in Lecture Notes In Computer Science, pages 370–381. Springer-Verlag, July 2001.

[4] 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.

[5] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley, New York, 1991.

[6] J. Desharnais, A. Edalat, and P. Panangaden. A logical characterization of bisimulation for labeled Markov processes. In *proceedings of the 13th IEEE Symposium On Logic In Computer Science, Indianapolis*, pages 478–489. IEEE Press, June 1998.

[7] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labeled Markov processes. In *Proceedings of CONCUR99*, Lecture Notes in Computer Science. Springer-Verlag, 1999.

[8] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Approximation of labeled markov processes. In *Proceedings of the Fifteenth Annual IEEE Symposium On Logic In Computer Science*, pages 95–106. IEEE Computer Society Press, June 2000.

[9] G. A. Edgar. *Integral, Probability and Fractal Measures*. Springer-Verlag, 1998.

[10] V. Gupta, T. A. Henzinger, and R. Jagadeesan. Robust timed automata. In O. Maler, editor, *Hybrid and Real-Time Systems*, volume 1201 of *Lecture Notes In Computer Science*, pages 331–345. Springer Verlag, March 1997.

[11] H. Hansson and B. Jonsson. A calculus for communicating systems with time and probabilities. In *Proceedings of the 11th IEEE Real-Time Systems Symposium*, pages 278–287. IEEE Computer Society Press, 1990.

[12] H. A. Hansson. *Time and Probability in Formal Design of Distributed Systems*, volume 1 of *Real-time Safety-critical Systems*. Elseiver, 1994.

[13] S. Hart and M. Sharir. Probabilistic propositional temporal logics. *Information and Control*, 70:97–155, 1986.

[14] J. E. Hutchinson. Fractals and self-similarity. *Indiana Univ. Math. J.*, 30:713–747, 1981.

[15] B. Jonsson and W. Yi. Compositional testing preorders for probabilistic processes. In *Proceedings of the 10th Annual IEEE Symposium On Logic In Computer Science*, pages 431–441, 1995.

[16] 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.

[17] D. Kozen. Semantics of probabilistic programs. *Journal of Computer and Systems Sciences*, 22:328–350, 1981.

[18] D. Kozen. A probabilistic PDL. *Journal of Computer and Systems Sciences*, 30(2):162–178, 1985.

[19] K. G. Larsen and A. Skou. Bisimulation through probablistic testing. *Information and Computation*, 94:1–28, 1991.

[20] P. D. 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.

[21] A. Philippou, I. Lee, and O. Sokolsky. Weal bisimulation for probabilistic processes. In C. Palamidessi, editor, *Proceedings of CONCUR 2000*, number 1877 in Lecture Notes In Computer Science, pages 334–349. Springer-Verlag, 2000.

[22] R. Segala. *Modeling and Verification of Randomized Distributed Real-Time Systems*. PhD thesis, MIT, Dept. of Electrical Engineering and Computer Science, 1995. Also appears as technical report MIT/LCS/TR-676.

[23] I. Ulidowski. Equivalences on observable processes. In *Proceedings of the Seventh IEEE Symposium On Logic In Computer Science*, pages 148–159. IEEE Press, 1992.

[24] I. Ulidowski. *Local Testing and Implementable Concurrent Processes*. PhD thesis, Imperial College, 1994.

[25] F. van Breugel and J. Worrell. An algorithm for quantitative verification of probabilistic systems. In K. G. Larsen and M. Nielsen, editors, *Proceedings of the Twelfth International Conference on Concurrency Theory - CONCUR'01*, number 2154 in Lecture Notes In Computer Science, pages 336–350. Springer-Verlag, 2001.

[26] F. van Breugel and J. Worrell. Towards quantitative verification of probabilistic systems. In *Proceedings of the Twenty-eighth International Colloquium on Automata, Languages and Programming*. Springer-Verlag, July 2001.

[27] S.-H. Wu, S. Smolka, and E. W. Stark. Composition and behaviors for probabilistic I/O automata. *Theoretical Computer Science*, 176(1–2):1–36, April 1997.