

RAPPORT DE SOUTENANCE

Nicolas Basset nous a présenté avec beaucoup de pédagogie une sélection du large éventail de résultats démontrés pendant sa thèse dans le domaine des automates temporisés.

Nicolas Basset pose les fondations d'une théorie nouvelle et pertinente abordée sous les angles variés de la dynamique symbolique, de la théorie de l'information et de la combinatoire. Le jury s'associe avec les rapporteurs pour souligner la grande qualité et l'originalité des résultats obtenus par N. Basset. Il a eu, par ailleurs, répondu avec vivacité et pertinence aux nombreuses questions du jury.

Nicolas Basset est d'ors et déjà un très bon chercheur, créatif et maîtrisant une large variété de domaines.

Le jury, unanime, lui décerne le titre de Docteur en Informatique de l'Université Marne-la-Vallée.

Champs-sur-Marne, le 5 décembre 2013

A. Bous
Arap
D. Jm
D. Ph

Les Membres du Jury

JEAN PAIRESSSE
Le Président du Jury

L'établissement délivre le titre de docteur sans mention ni félicitations du jury

SOUTENANCE DE THESE


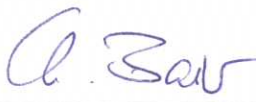

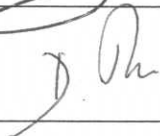
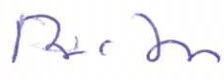
du Jeudi 05 décembre 2013

LISTE D'EMARGEMENT DES MEMBRES DU JURY

Nom et Prénom du Candidat : BASSET Nicolas

Titre de la Thèse : Volumétrie des langages temporisés et applications.

Nom, prénom du Président du Jury : MAIRESSE JEAN

Nom, Prénom des membres du Jury	Emargement
ALUR Rajeev	
ASARIN Eugène	
BAIER Christel	
MAIRESSE Jean	
PERRIN Dominique	
VALLÉE Brigitte	

Rapport de soutenance (voir page suivante)

UNIVERSITY of PENNSYLVANIA

School of Engineering and Applied Science
Department of Computer and Information Science
Levine Hall
3330 Walnut Street
Philadelphia, PA 19104-6389
Telephone: (215) 573-7483
Email: alur@cis.upenn.edu

November 18, 2013

Dissertation Committee
Universite Paris

Dear Colleagues:

Enclosed is my report on the thesis titled *Volumetry of timed languages and applications* submitted by Nicolas Basset. The results in the thesis are of high quality, and I am happy to recommend its acceptance.

The object of study in this thesis is the structure of timed languages, that is, sets of sequences of events along with real-valued delays corresponding to the times of occurrences of events. Timed automata, finite-state generators of such timed languages, are useful in modeling and analysis of embedded systems, and have been extensively studied for over twenty years now. This thesis brings fresh insights and new perspectives to this study. The starting point for this study is the notion of *volumetric entropy* of a timed language that formally captures the rate of growth of the number of words in the language with respect to the length of the words. This notion of entropy was first introduced by Asarin and Degorre a few years ago, and Basset's thesis presents new results and explores new connections.

Chapter 3 shows how the notion of entropy can be useful in understanding the notion of *robustness* for timed automata. Intuitively, the behavior of a timed automaton is pathological if realizing it requires clocks with infinite precision. There have been past attempts to formalize robustness, but Basset's formalization using *thin* and *thick* automata is very compelling due to the connection he establishes with the entropy. The results relating this classification to the existence of the so-called *forgetful* cycles provide useful insights.

Chapter 4 establishes the existence of a stochastic process of maximal entropy corresponding to a timed automaton. It shows how to construct such a process, and proves a number of interesting properties such as ergodicity. These results lift classical results on finite graphs and Markov chains to the timed setting, and have potential application to uniform sampling of runs of timed systems,

and thus to quantitative analysis.

Chapter 5 connects the theory of timed languages to symbolic dynamics. This requires the design of an appropriate shift operation on infinite alphabets. These tools also lead to an interesting result that connects the volumetric entropy to the entropy of the discrete process obtained by discretization, in the limit.

Chapter 6 develops timed analogs of classical theory of channel coding for discrete messages; and Chapter 7 exactly characterizes the sequence of volumes of timed words using generating functions. Finally, Chapter 8 shows an application of the characterization of the volumes of timed languages to combinatorial problems related to counting and generating permutations and uniform sampling.

The results in the thesis are developed with mathematical rigor, and draw upon tools from a number of mathematical subdisciplines. The technical difficulty of some of these results is high as a clever combination of combinatorial and analytical arguments is needed. Each of the chapters 3, 4, 5, 6, 7, and 8 contain results that can be published as papers on their own in premier conferences (indeed many of these have already appeared as publications).

In summary, both in terms of new insights and technical depth, I consider the thesis to be exceptional, and recommend its acceptance with enthusiasm. Please do not hesitate to contact me if you need more information.

Best regards,

Rajeev Alur
Zisman Family Professor of Computer
and Information Science

Brigitte VALLÉE,
Directrice de Recherche au CNRS,
brigitte.vallee@unicaen.fr

Rapport sur le manuscrit de thèse de Nicolas Basset

Volumetry of timed languages and applications.

Contexte général. Il s’agit d’une thèse d’informatique mathématique qui se situe au carrefour de plusieurs domaines : systèmes dynamiques discrets (automates, théorie des langages et dynamique symbolique) vérification, problèmes de comptage et de génération uniforme.

La motivation générale est apportée par la vérification, où les automates classiques et les langages réguliers ne sont pas complètement adéquats, car construits de manière fondamentale sur un temps discret. Il faut leur ajouter une notion de temps continu, qui permet de formaliser, puis de quantifier le délai entre deux événements (deux transitions successives de l’automate, deux lettres successives d’un mot du langage). C’est ainsi que la notion d’automate temporisé a été introduite par Alur et Dill dans les années 90. C’est devenu depuis un modèle très utilisé en vérification.

Mais les fondements théoriques de ces automates temporisés ne sont pas encore complètement bien établis. Ce n’est qu’en 2009 que Eugène Asarin et Aldric Degorre se sont penchés sur les propriétés quantitatives de ces automates temporisés, et de leurs langages associés, appelés langages temporisés. Ces versions temporisées (automates, langages) peuvent être vues comme des extensions continues de leurs analogues classiques. Les cardinaux des études classiques (nombre de mots de longueur n , nombre de chemins du graphe de longueur n) sont remplacés par les volumes de polytopes associés aux contraintes de temps portant sur les mots de longueur n , dans le cadre temporisé. L’étude quantitative des modèles temporisés est centrée sur l’étude asymptotique des volumes quand la taille n tend vers l’infini. C’est cette asymptotique qui permet par exemple de définir l’analogie de l’entropie classique, qu’on appelle entropie volumétrique, désignée par \mathcal{H} dans la suite.

Description des principaux résultats de la thèse. La thèse de Nicolas Basset se place dans ce cadre dessiné par Asarin et Degorre, le re-visite et le prolonge. C’est une thèse qui contribue fortement à établir les fondements du modèle temporisé. NB caractérise d’abord (dans le chapitre 3) les pathologies du modèle, pour les écarter de son étude ultérieure (il y reviendra seulement dans le chapitre 5), puis il se concentre sur le cadre ”temporisé non-pathologique”, et montre que, sur des aspects très divers, on peut définir des extensions des outils “classiques” qui permettent de répondre finement à des questions “naturelles” dans le cadre temporisé.

Chapitre 3. C’est un chapitre issu d’un article écrit avec Asarin, présenté à la conférence Formats11. Il s’agit d’opérer une classification des automates temporisés, pour distinguer ceux qui ont un comportement naturel – les “gras” –, de ceux qui sont pathologiques – les “maigres” –. La terminologie adoptée provient de la volumétrie, car la classification opère sur la dimension des polytopes sous-jacents, qui est “pleine” dans le cas gras, et dégénère dans le cas maigre. Je n’en dirai pas plus sur ce chapitre, qui utilise des techniques difficiles et variées – outils spécialisés au domaine “temporisé” mêlés ‘a des arguments à la Ramsey–domaines où je ne me sens pas compétente.

Chapitre 4. C’est un chapitre issu d’un article que NB a écrit seul, et qui a reçu à ICALP 2013 la récompense du meilleur article étudiant (Track B). Asarin et Degorre ont montré

qu'on pouvait remplacer, dans le cadre temporisé, la matrice de transition P de l'automate classique par une matrice Ψ d'opérateurs intégraux. Ils ont aussi expliqué le rôle que les "objets spectraux dominants" (valeur propre dominante, fonction propre dominante) de cet opérateur (et de son adjoint) jouent dans la détermination des propriétés quantitatives de l'automate. Ils ont ainsi étendu aux automates temporisés ce qu'on sait faire pour les automates classiques, où ce sont les objets spectraux dominants de la matrice P et de son adjointe qui jouent un tel rôle.

Dans sa thèse, Nicolas Basset re-visite les résultats de Asarin et Degorre, avec des hypothèses moins restrictives et utilise un espace fonctionnel différent, de type espace de Hilbert, où l'opérateur Ψ devient un opérateur matriciel dont les composantes sont des opérateurs intégraux à noyau. C'est donc un opérateur compact. Le cadre est ainsi plus naturel, et on saisit mieux l'analogie avec le calcul matriciel.

Il s'intéresse aussi à des propriétés d'entropie maximale; il montre qu'elles sont reliées, aux objets spectraux dominants de l'opérateur Ψ , comme c'est le cas des chaînes de Markov sur un graphe classique, où les propriétés similaires sont reliées, aux objets spectraux dominants de la matrice P . Comme dans le cas classique, NB explique qu'elles conduisent à un choix optimal des probabilités de transition, qui permettra, si on la choisit en entrée, une distribution des calculs "quasi-uniforme" dans l'automate temporisé

Il est souvent fait allusion à de nouveaux résultats sur le saut spectral de Ψ , présents dans l'article [ADB13]. Il est dommage que ces résultats ne soient pas un tout petit peu décrits, et que, de manière générale, la liaison entre le saut spectral, la décomposition spectrale de Ψ et l'existence de termes de reste exponentiellement petits dans l'expression asymptotique des volumes, ne soient pas davantage mis en lumière, dans ce chapitre comme dans le chapitre 7, par exemple.

Chapitre 5. C'est un chapitre qui définit des éléments de dynamique symbolique temporisée, et qui introduit aussi deux autres mesures. La première mesure est l'entropie volumétrique discrétisée associée à un pas $\epsilon > 0$, et désignée par h_ϵ . C'est l'analogie de l'entropie volumétrique quand le temps devient discret et ne prend ses valeurs que dans l'ensemble $\mathbb{N}\epsilon$; l'automate temporisé redevient alors un objet discret, la dynamique symbolique reliée travaille sur des espaces compacts, et rend le problème discrétisé plus facile à aborder. Avec le point de vue de la théorie de l'information, il est assez naturel de comparer la différence $h_\epsilon - |\log_2 \epsilon|$ avec l'entropie \mathcal{H} et de se poser la question de la convergence du premier terme vers le second quand le pas ϵ tend vers 0. C'était une question ouverte, posée par Asarin et Degorre, et NB la résout dans ce chapitre. Il me semble qu'il obtient aussi un terme de reste explicite (même s'il ne le dit pas vraiment), qui s'avère important si l'on veut utiliser ce résultat pour calculer \mathcal{H} à partir de son analogue discret h_ϵ .

La deuxième "mesure" est une dimension, la dimension métrique moyenne, notée **mdim**, et introduite par Gromov. C'est une sorte de dimension de Hausdorff, qui quantifie la "dégénérescence" de l'automate au sens du Chapitre 3. En effet, un automate dégénéré produit des polytopes qui ne sont pas toujours de dimension pleine, et NB montre que c'est ce phénomène qui est quantifié par la dimension **mdim**.

Chapitre 6. C'est un chapitre issu d'un article écrit avec Asarin, Béal, Degorre et Perrin, présenté à la conférence MFCS2012. Il s'agit ici de re-visiter des problèmes classiques de théorie des codes, cette fois-ci dans le cas des langages temporisés. Il y a trois cas distincts selon que le canal reste discret, la source ou le canal deviennent tous deux temporisés, avec la même échelle de temps ou non (avec la même notion de temporisation discrète que celle qui est définie au chapitre précédent). Les résultats obtenus dans ce chapitre correspondent à une généralisation naturelle du cadre classique – mais naturelle ne signifie pas facile!... Je n'en dirai pas plus sur ce chapitre.

Chapitre 7. C'est un chapitre issu d'un article que NB a écrit avec Asarin, Degorre et Perrin. Il adopte un point de vue combinatoire, et introduit le point de vue de la combinatoire "volumétrique", qui est l'analogie pour les automates temporisés de la combinatoire énumérative classique. Dans le cadre classique, la série génératrice (ordinaire) $f(z)$ d'un

langage régulier s'exprime en fonction du quasi-inverse $(I - zP)^{-1}$ qui fait intervenir la matrice de transition P de l'automate associé. Ici, les auteurs montrent que le parallèle établi au chapitre 4 se poursuit : la série génératrice d'un langage temporisé s'exprime en fonction du quasi-inverse $(I - z\Psi)^{-1}$ qui fait intervenir l'opérateur Ψ du chapitre 4. Les auteurs se concentrent surtout sur les cas où les calculs permettent d'obtenir une formule close. On pourrait aussi faire de l'asymptotique... (voir remarque ci-dessus).

Chapitre 8. C'est un chapitre issu d'un article écrit par Nicolas seul, actuellement soumis. Il s'agit ici d'étudier une question de combinatoire énumérative (et analytique), dont les motivations sont donc extérieures au cadre temporisé. On considère l'ensemble des permutations dont la signature appartient à un langage régulier, $L \subset \{a, d\}^*$, où a signifie ascent (montée) et d signifie descente. L'étude de telles permutations (via leur série génératrice exponentielle) est un vieux problème classique, mais qui n'a pas été étudié sous cette forme très générale. Et la génération aléatoire de telles permutations est un sujet plus récent, là non plus, pas étudié dans cette généralité.

Bien sûr, il y a du temporisé qui s'introduit, et c'est là toute l'originalité de l'approche. NB travaille avec des langages temporisés qui sont des extensions du problème de départ, et qui sont sur l'alphabet $\{S, T\}$ avec la signification $S = \text{straight}$, et $T = \text{turn}$. En écrivant le système des équations "à la Kleene" de l'automate temporisé, dans l'esprit du chapitre précédent, il obtient une expression explicite de la série génératrice exponentielle recherchée. Il obtient aussi avec ce point de vue temporisé un algorithme de génération uniforme, qui étant donné le langage L et une longueur n , engendre une permutation aléatoire de longueur n dont la signature appartient à L .

Évaluation. C'est une thèse au spectre très large, qui explore "sous toutes les coutures" un domaine encore peu exploré – celui des langages temporisés – et qui contribue notablement à le fonder, en se posant toutes les questions naturelles auxquelles on peut songer : " Et si maintenant j'ai un langage temporisé et non plus un langage classique ? puis-je faire ce que je savais faire avant ? puis-je coder ? mesurer ? calculer ? " Le même objet est ainsi exploré avec une multitude de points de vue complémentaires, et la thèse jouit donc d'une unité profonde. Nicolas Basset démontre ici une culture étendue, des compétences très larges, la maîtrise de méthodes et techniques très diverses. Il faut signaler que ces résultats ont déjà fait l'objet de publications dans de très bonnes –voire excellentes– conférences internationales (deux comme seul auteur). Il lui faut vraiment songer à un long article de revue qui reprendrait le point de vue de synthèse adopté par la thèse.

Il y a bien sûr quelques petits bémols. La thèse est souvent bien écrite, avec un anglais un peu approximatif parfois, et des phrases à rallonges, à la ponctuation souvent parcimonieuse. L'organisation est claire "dans son gros grain", mais l'est moins dans son grain plus fin. Il y a beaucoup d'abréviations et un petit lexique serait alors bien utile ; le lecteur doit avoir tout retenu dès la première fois, car, après, on ne l'aide pas, et il y a peu de renvois ; l'architecture des preuves, parfois complexes, n'est guère explicitée ; il y a un manque d'explications informelles, et les exemples, trop rares, sont insuffisamment utilisés.

Les motivations –en particulier celles liées au domaine de la vérification – ne sont pas suffisamment explicitées ; elles le sont davantage dans les articles liés à la thèse. Et le "retour sur investissement" est parfois aussi un peu déficient : "le résultat que je viens d'obtenir s'applique-t-il à une situation "réelle" ? laquelle ? ". C'est un peu dommage, car cela donnerait encore davantage de relief aux excellents résultats obtenus.

Tout ce qui précède montre sans aucune ambiguïté que je suis très favorable à la soutenance de cette très bonne thèse, excellente à beaucoup d'égards.

Le 11 novembre 2013,

Brigitte VALLÉE

