Report on the PhD Thesis of Gaëtan Staquet

Promotors: Prof. Véronique Bruyère (UMONS) and Prof. Guillermo Alberto Pérez (UAntwerpen).

Jury members: Prof. Dana Fisman (Ben-Gurion University), Prof. Daniel Neider (TU Dortmund University), Dr. Ocan Sankur (CNRS, Université de Rennes), Prof. Frits Willem Vaandrager (Radboud Universiteit), and Prof. Jef Wijsen (UMONS).

Gaëtan Staquet defended privately his thesis, entitled *"Active Learning of Automata with Resources"*, on June 19, 2024, at UAntwerpen, in front of the jury except Prof. Dana Fisman who attended remotely. After this defense, the jury decided that the thesis could be defended publicly. The public defense took place on September 11, 2024, at UMONS.

The general topic of the thesis is active automata learning. The main contributions are described below.

Active automata learning. In order to understand and verify complex systems, we need accurate models that are either understandable for humans or can be analyzed fully automatically. Such models, however, are typically not available for legacy software and for AI systems constructed from training data. Model learning is a technology that potentially may fill this gap. Model learning, also known as active automata learning, is a black-box technique for constructing state machine models of software and hardware components from information obtained through testing (i.e., providing inputs and observing the resulting outputs). It has been successfully used in numerous applications, for instance for spotting bugs in implementations of several major network protocols.
Active learning is often abstracted as an interaction between a learner trying to infer a model of an unknown machine and a teacher that can answer queries. Angluin's seminal L* algorithm implements a learner for finite-state automata while using membership queries (Is a word accepted by the teacher automaton?) and equivalence queries (Is the learned automaton equivalent to the teacher automaton?) assumed to be answered by the teacher.
Though recent works have greatly advanced the state of the art in finite-state automata learning, handling real-world applications usually involves tailor-made abstractions to circumvent elements of the system which result in an infinite state space. This highlights the need for learning algorithms that focus on more expressive models. The aim of the thesis is to investigate three classes of automata with additional resources: realtime one-counter automata, visibly-pushdown automata (VPAs), and automata with timers. For the first and third classes, new active learning algorithms are provided. Concerning the second class, a new streaming validation algorithm is proposed for JSON documents, partially based on a known learning algorithm for VPAs.

Learning realtime one-counter automata. The first studied model of automata enriched with resources is that of one-counter automata. These are automata extended with a natural counter that can be incremented/decremented as well as tested against zero along transitions. As the counter value is generally not bounded, this model allows to capture the behavior of some infinite-state systems. It is known to be expressive enough to verify programs with lists or to validate streams of certain classes of XML documents.
The thesis presents a new learning algorithm for a subclass of one-counter automata called real-timed. Even though we are dealing with infinite-state systems, their behavior can be described finitely and learned using Angluin's approach, from which a realtime one-counter automaton can be extracted. Nevertheless, this requires non-trivial adaptations using two additional types of queries including a counter-value query (the teacher is assumed to have access to an executable black box with observable counter values). Additionally, significant effort is needed to prove that the algorithm always terminates. The proposed algorithm runs in exponential time and space and uses an exponential number of queries in the worst case.
This learning algorithm has been implemented and evaluated on random realtime one-counter automata, as well as in a case study involving the validation of JSON documents.

Validating JSON documents. JavaScript Object Notation (JSON) has overtaken XML as the de facto standard for data exchange, particularly for web applications, due to its ease of readability by both humans and computers. In a JSON document, objects are unordered collections of key-value pairs, while arrays are ordered collections of values. JSON Schema is a simple schema language that allows users to impose format constraints on the structure of JSON documents.
The second part of the thesis investigates the streaming validation of JSON documents against JSON schemas. This means that the document is received piece by piece (for instance, when transmitted over a network), and the task is to check whether it satisfies the schema. Classical validation algorithms are not efficient in the streaming context as they need to know the entire document before validation. The new approach proposed in the thesis uses a visibly-pushdown automaton (VPA), representing the JSON schema, that reads a given JSON document while validating it. As the VPA assumes a fixed order on the key-value pairs of the objects (in order to keep its size reasonable), the streaming algorithm requires an additional graph to efficiently manage the permutations of the key-value pairs.
Prior to developing this algorithm, the thesis establishes that VPAs are expressive enough to capture the language of JSON documents satisfying any JSON schema. Moreover, it explains that active learning of VPAs is a good alternative to the automatic construction of such a VPA from a given JSON schema.
Finally, the new validation algorithm has been implemented and evaluated on various JSON languages. Experimental results indicate that this algorithm exhibits good performance.

Mealy machines with timers. The last part of the thesis studies automata that use timers to encode timing constraints. A timer is initialized with a certain value and decreases over time. When it reaches zero (indicating a timeout), a special event occurs that must be handled, similar to how interruptions are managed in a processor. Automata with timers constitute a strict subclass of timed automata. Before investigating active learning of this model, the thesis first proves that the configuration reachability problem for automata with timers is PSPACE-complete, similar to the case for timed automata. It then examines their non-deterministic behavior in situations where multiple timers time out simultaneously, or when the user provides input at the exact moment a timer times out. Race-avoiding automata with timers are designed to avoid such non-deterministic behaviors. The thesis provides an effective characterization of race-avoiding automata with timers and proves that the related decision problem is in 3EXP and PSPACE-hard.
Finally, the thesis introduces an active learning algorithm for Mealy machines with timers that are race-avoiding. This algorithm is a non-trivial extension of the L# algorithm of Vaandrager et al. to a timed setting. It requires a number of queries that is factorial in the number of timers and polynomial in the number of states. In practice, the number of timers remains relatively small and can even be fixed a priori, which leads to a polynomial algorithm. Experiments with a prototype implementation show that this learning algorithm can efficiently learn realistic benchmarks.

Obtained results. Most of the results presented in this thesis have undergone peer review and have been published in selective international conferences, including TACAS '22, TACAS '23, and FORMATS '23. In the collaborative work that led to these publications, Gaëtan Staquet played a key role in developing the theory, implementing it, and validating it experimentally, thus significantly shaping the direction and success of the collaboration. The jury has determined that the obtained results are novel, deep, technically challenging, and numerous. Additionally, the jury underscores the importance of implementing the algorithms, rather than solely developing theoretical proofs of their correctness. All developed tools are open source and publicly accessible. The thesis concludes with several compelling open questions that address both theoretical and practical aspects of active learning for models with resources.

Quality of the document. The high quality of the document is unanimously recognized by the jury. The inclusion of numerous illustrative examples is also highly valued by all members. Given the outstanding quality of the written work, no changes are required for the final version, aside from correcting a few minor typos.

Quality of the presentation and answers to the questions (private defense). Gaëtan Staquet presented the results of his thesis in a clear and pedagogical manner. He provided an overview of

the main theorems, as well as several interesting technical challenges encountered during his research. The jury unanimously appreciated his presentation. All members posed both technical and general questions, including some about future research directions. Gaëtan answered each question clearly and demonstrated a level of maturity and enthusiasm for research that convinced the jury he is well-positioned for a successful career in academia.

Quality of the presentation and answer to the questions (public defense). At the public defense, Gaëtan Staquet delivered a highly pedagogical presentation tailored to a broader audience. He effectively summarized his main contributions and explained the potential impact of his research topic. The questions varied in scope, ranging from broad, overarching topics to specific technical details. Gaëtan Staquet responded to each question thoughtfully without resorting to speculation.

Conclusion. Gaëtan Staquet's results are profound and significantly advance the scientific understanding of active automata learning and its applications. The quality of his PhD thesis manuscript is exceptional. The jury also greatly appreciated the quality of his oral presentations during both the private and public defenses. The software tools developed during his thesis are expected to become a valuable baseline for the community in learning (legacy) models. During the question session, the candidate demonstrated thorough mastery of all the results presented in his thesis. As a result, the jury unanimously decides to award Gaëtan Staquet the title of "Docteur en sciences".

Le président (Jef Wijsen)

Le secrétaire (Ocan Sankur)

Véronique Bruyère
UMONS, Avenue Maistriau 15, 7000-Mons, Belgium,
Guillermo A. Pérez
UAntwerp, Middelheimlaan 1, 2020-Antwerp, Belgium

March 21, 2025

Hiring committee, Bordeaux INP, MTV

Dear Madam, Sir,

As the two PhD supervisors of Gaëtan Staquet at the Universities of Mons (UMONS) and Antwerp (UAntwerp), we recommend him most highly and without any hesitation for a position as *maître de conférences* at your institution. Our reasoning is outlined below.

We have had the pleasure of witnessing Gaëtan's growth as a researcher and his contributions to the field of computer-aided verification and in particular to topics around model learning. Gaëtan has a master's degree in computer science from UMONS. During his studies, he did a research internship at UAntwerp with Guillermo A. Pérez's team. At the same time, he also coauthored a paper with Jef Wijsen (UMONS) on a different database-related topic.

Thereafter, Gaëtan secured personal funding for his PhD in the form of a "Fonds de la Recherche Scientifique" (FNRS) scholarship, which is very selective. He obtained, in September 2024, a PhD at UMONS and UAntwerp under our joint supervision. He did so within 4 years, the usual duration of a PhD in Belgium. Gaëtan is currently an Inria postdoctoral researcher in Rennes University, with Nathalie Bertrand, head of DEVINE team.

Throughout his graduate studies, and during his research positions, Gaëtan showed versatility in adapting and learning about new topics ranging from theory of databases to automata and game theory. Within automata theory, his PhD thesis deals with various models having access to different forms of "memory" such as counters, stacks, and clocks. His current postdoctoral work focuses on parametrized games played on graphs.

Gaëtan is one of the promising young researchers focusing on applied active learning of timed models. This is supported by the best paper award he won for his work on the topic presented at the 21st International Conference on Formal Modeling and Analysis of Timed Systems. In addition, he spent three months as a visiting researcher in Radboud University, The Netherlands, under the guidance of Frits W. Vaandrager. The composition of the jury for his PhD defense is also telling of the level and topics of the work of Gaëtan: It included world experts on model learning Dana Fisman (Ben-Gurion University, Israel), Daniel Neider (TU Dortmund, Germany), and Frits W. Vaandrager.

During his PhD, Gaëtan played a key role in developing the theory, implementing it, and validating it experimentally. His PhD jury determined that the obtained results were novel, deep, technically challenging, and numerous. In particular, Gaëtan successfully developed new learning algorithms for three models of automata with memory that are far from trivial. Moreover, he has a talent for translating these algorithms into practical prototype implementations, enabling empirical evaluation of algorithms. It is worth mentioning that these are well-engineered prototypes with

high-coverage testing, designed and programmed following excellent software engineering practices. In fact, Gaëtan participated (and ranked high) in more than one coding event during his graduate studies.

Besides research, Gaëtan actively participated in educational tasks during his time at UMONS but also now as a postdoctoral researcher in Rennes. He has helped with courses on algorithmic and experimental complexity, logic programming, functional programming, and has supervised more than one masters' thesis. His assistance with courses at UMONS was highly appreciated by the staff and the students. In addition, he is motivated to popularize his work. He even participated in a "my thesis in 180 seconds" competition organized in the French-speaking part of Belgium.

He has an impressive ability to articulate complex ideas clearly and concisely both in written and oral form. He has had many opportunities to exercise this since he was invited to give presentations on his work, for instance, at seminars hosted by EPITA in Paris, LS2N in Nantes and, and University of Bordeaux. He also willingly presented his work during some of the summer schools and workshops he attended since starting his PhD (see `https://www.gaetanstaquet.com/academic/gaetan_staquet.pdf` for the full list).

Finally, during the last two years of his PhD, Gaëtan actively built a strong network of collaborations within France and the Netherlands, demonstrating his initiative and leadership potential.

For all the reasons above, we wholeheartedly recommend Gaëtan Staquet for the position of *maître de conférences*.

Please do not hesitate to contact us if you have any further questions.

Sincerely,

V. Bruyère and G. A. Pérez

Radboud Universiteit

Tho who it may concern

Faculty of Science
Institute for Computing and Information Sciences

Datum 27 maart 2025
Ons kenmerk -
Referentie -
Betreft **Letter of Recommendation for Gaetan Staquet**

Dear Sir/Madam,

This letter is a strong recommendation for Gaëtan Staquet!

I know Gaëtan and his work very well. He visited my group at Radboud University from August to October 2023, and during this period we interacted on a daily basis. Based on this visit, we wrote two articles together: one on Mealy machines with timers, published at FORMATS'23 (where it received a best paper award), and one on learning of Mealy machines with timers, under preparation for journal submission (with a first version published on arXiv). I was member of the jury for Gaëtan's PhD defense both at the Universiteit van Antwerpen and at the Université de Mons.

The main focus of my current research is the theory and application of active automata learning, which is a black-box technique for constructing state machine models of software and hardware from information obtained through testing (i.e., providing inputs and observing the resulting outputs). Automata learning has been successfully used in many applications, e.g., for spotting critical bugs in implementations of major network protocols such as TCP, SSH, TLS and DTLS. Timing plays a crucial role in these network protocols. However, extending model learning algorithms to a setting that incorporates quantitative timing information turns out to be challenging. Twenty years ago, the first papers on this subject were published, but we still do not have scalable algorithms. However, largely due to Gaëtan's work on learning of Mealy machines with timers (jointly with me, my student Bharat Garhewal, and Gaëtan's thesis supervisors Véronique Bruyère and Guillermo Pérez), timing-aware learning algorithms that scale to realistic network protocols is now within reach. Such algorithms will have numerous applications, not only for spotting

Mercator building
Toernooiveld 212
6525 EC Nijmegen
+31 24 365 2216

Postbus 9010
6500 GL Nijmegen
The Netherlands

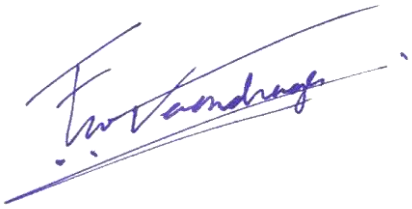
https://www.cs.ru.nl/F.Vaandrager/

bugs, but e.g. also for refactoring of legacy software. The results obtained by Gaëtan are deep and significantly extend the scientific knowledge about active automata learning.

Within our research team working on learning of timed systems, Gaëtan is the central person: he is the one who always keeps the overview but is also aware of all the nitty gritty details, both in the mathematical theory and in the software. Frequently, when one of us had an idea, he would work out the details, explain to us that the original idea was flawed, but that he had found a modification that worked. Gaëtan's strengths as a researcher are his passion for and deep/broad knowledge of theoretical computer science, his creativity, his discipline to work out all the details, his ability to perform research both individually and as part of a team, and his talent to clearly explain results to his colleagues. Like me, all members of the PhD jury praised the remarkable quality of the thesis. His clear answers to our questions demonstrated a level of maturity and passion for research, and convinced all of us that he will have successful career in academia.

Any research institute should be thrilled to have Gaëtan Staquet amongst its staff! Please do not hesitate to contact me in case you have further questions.

Yours sincerely,



Prof.dr. Frits Vaandrager
Head Department of Software Science
T +31 24 365 2216
E F.Vaandrager@cs.ru.nl
W https://www.cs.ru.nl/F.Vaandrager/

UMONS

Université de Mons

Lettre de recommandation concernant
les activités d'enseignement de Gaëtan Staquet, assurés en
Faculté des Sciences de l'Université de Mons (Belgique)

Mons, le 11 mars 2025

Monsieur / Madame,

Je suis professeur ordinaire, directeur du laboratoire de génie logiciel et responsable de plusieurs cours liés à la programmation et développement logiciel. Par cette lettre, je souhaite vous faire part de mon opinion concernant les activités d'enseignement de Gaëtan Staquet, pendant qu'il était chercheur doctorant dans le Département de Sciences Informatiques de l'Université de Mons (UMONS).

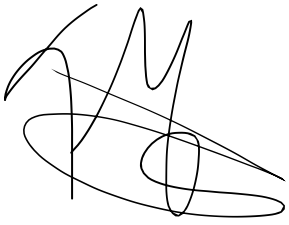
J'ai eu l'honneur d'avoir Gaëtan Staquet comme assistant pour les séances de travaux dirigés (TD) pour mes cours de « programmation fonctionnelle » et « programmation logique » dispensés en troisième année du bachelier en sciences informatiques en Faculté des Sciences. (Pour rappel, le diplôme de master en sciences informatique en Belgique nécessite 5 années d'études correspondant à 300 crédits ECTS, dont trois années de bachelier et deux années de master.) L'implication de Gaëtan correspondait à 30 heures de TD par année académique (14 pour programmation fonctionnelle et 16 pour programmation logique), et incluait également la création et correction des questions d'examens, ainsi que la préparation de l'énoncé des TD.

Je suis extrêmement content de l'implication de Gaëtan, car il s'est toujours acquitté de ces tâches avec dévouement et rigueur. Je suis donc convaincu que Gaëtan sera un enseignant excellent pour des cours, projets ou travaux dans le domaine large de l'informatique, que ce soit pratique ou théorique.

Je vous prie d'agréer l'expression de mes sentiments distingués.

Prof. Dr. Tom Mens
Département de Sciences Informatique
Faculté des Sciences, Université de Mons

tom.mens@umons.ac.be
https://informatique.umons.ac.be



Thomas Genet
ISTIC/IRISA
Campus de Beaulieu
35042 Rennes cedex

Recommandation de Gaëtan Staquet
pour un poste de maître de conférences

Rennes, le 10 mars 2025

A qui de droit,

J'écris cette lettre pour soutenir la candidature de Gaëtan Staquet pour un poste de maître de conférences. Gaëtan a enseigné en 2024-2025 dans l'U.E. de L1 ISTN intitulée "Algorithmique et complexité expérimentale" (`http://people.irisa.fr/Thomas.Genet/ACX/`), dont je suis responsable. Il a encadré 28,5 heures de travaux pratiques pour un groupe de 20 étudiants. Gérer un groupe de travaux pratiques pour cette U.E. de L1 demande beaucoup d'implication et d'organisation aux encadrants. En particulier, les encadrants sont chargés de réaliser l'évaluation de leur étudiants par des entretiens individuels en temps limité. Gaëtan s'est acquitté de ces tâches avec enthousiasme et a été d'un grand soutien en se chargeant de la relecture et correction des sujets de TP et des sujets d'examen terminal. J'ajoute que Gaëtan est une personne avec qui il est facile d'interagir et de travailler.

Tout ceci fait que je soutiens sans réserve la candidature de Gaëtan Staquet pour un poste de maître de conférences. Je suis sûr que son intégration sera profitable à toute l'équipe pédagogique de l'UFR dans lequel il s'établira.

Thomas Genet
Professeur des universités
ISTIC/IRISA, Université de Rennes