Model Checking Cancer Automata

Juliana K. F. Bowles and Agastya Silvina
School of Computer Science, University of St Andrews, St Andrews, KY16 9SX, UK
{jkfb|as362}@st-andrews.ac.uk

***Abstract*— Cancer is a chronic disease where cells grow and multiply in an uncontrollable manner ultimately spreading and invading surrounding tissue, and metastasising in other parts or organs of the body. Automata can be used to capture cancer evolving through a (discrete finite) sequence of progressive stages called *phenotypes*. Automata consist of states (known as hallmarks of cancer) and transitions between states, indicating a progression or regression of the cancer. We explore extensions and combinations of different variants of timed automata and associated tools to model and analyse a model of the disease in different ways. We combine patient information and comorbidities with the cancer automaton through composition. The goal of this work is to use model checking as an analysis technique to provide further insights into the effectiveness of treatment plans for a given patient, and how these could potentially inhibit or slow down the progression of cancer.**

I. INTRODUCTION

Cancer is a progressive disease of abnormal cells which grow and multiply in an uncontrollable manner, ultimately spreading and invading surrounding tissue and metastasing in other parts or organs of the body [1].

Cancer's growth rate varies and can quickly make up to trillions of cells. Tumours are masses or groups of cells which appear on solid or liquid tissues and can, in principle, be either benign or malignant. A tumour is malignant if it consists of cancerogenous cells. Cancer is a malignant tumour and is dangerous as it can cause death if left untreated. When cancer grows it can spread and invade nearby tissues and metastasise.

There are many factors that influence the growth of cancer such as an unhealthy diet, smoking, physical inactivity, among others [2]. As the life expectancy of people increases, cancer also becomes more common. According to [3], about 589,430 Americans are expected to die of cancer in 2015 which amounts to about 1,620 people per day. It is the second most common cause of death in the US and is responsible for nearly 1 of every 4 deaths [3]. Similar statistics can be found in Europe (see[1] for the UK). Cancer research and effective forms of treatment are thus of upmost importance. Considerable research is done in improving the understanding of cancer progression. Amongst the efforts on researching cancer and cancer treatments are cancer modelling and tumour growth analysis.

Cancer modelling has been addressed in different fields ranging from applied mathematics to computer simulation. In particular, probability models (such as continuous Markov chains) are widely used to represent tumour growth.

In model checking, systems are modelled as finite state automata. Automata can be used to capture cancer evolving through a (discrete finite) sequence of progressive stages called *phenotypes*. Cancer Hybrid Automata (CHA) [4] have been introduced to model the progression of cancer with discrete phenotypes. It consists of states (aka stages or hallmarks of cancer) and state transitions indicate cancer progression or regression. Transitions are labelled by a treatment (seen as a set of drugs) which if given to the patient (in a particular dosage) inhibit the transition and hence the evolution of the cancer. In particular, adding medication and treatments, a regression of the overall cancer state may become possible. The idea of an approach such as CHA [4] is that through verification and analysis of cancer models we can clarify the effect of a treatment, simulate cancer evolution and ultimately explore ways (different medications and treatments) to avoid reaching metastasis. CHA is a conceptual framework based on realistic biological foundations, but lacks tool support and automated analysis.

The paper is structured as follows. In the next section we give a brief description of our contribution. In Section III, we describe an overview of our approach. Section IV briefly describes some of our models, strategies and model combinations, as well as analysis results. We conclude the paper with ideas for further work.

II. OUR CONTRIBUTION

The main motivation of our work is to find an alternative model suitable for capturing cancer progression which solves some of the problems that CHA [4] currently has. First, a cancer automaton must include all required information on possible treatments (for the particular type of cancer) as known in the medical literature and current practice. Secondly, the model has to be compositional and as such make it possible to consider additional patient information and *comorbidities*. Comorbidities relate to the presence of additional diseases other than the primary one which in our case is cancer [5]. Additional diseases (possibly chronic ones) cannot be ignored as these may affect certain organs in critical ways, affect the responsiveness of medications, and so on. Finally, tool support is essential, and from an analysis point of view, we focus on:

- How to obtain traces from a cancer automaton corresponding to *positive treatments*, that is, treatments which show evidence of the cancer being controlled

*This work was partially supported by EPSRC grant EP/M014290/1.

[1] www.cancerresearchuk.org

and possibly even reverted, taking into account any comorbidity.
- Property verification and model correctness. Properties are used to compare positive treatments and medications used by determining the time it takes to revert to an earlier cancer stage, possible costs or other metrics at hand.

The specified properties (written in timed and probabilistic variants of CTL [6]) can help us to compare the effectiveness of different treatment plans for a given patient, in terms of how they can inhibit the cancer not just to prevent it from reaching metastasis but respecting other concomitant conditions. Examples of properties include *it is possible that metastasis will never be reached* (which would correspond to a path where the treatment is effective), *the probability of cancer cells reaching metastasis in less than 5 years* (which gives a probability value for that scenario), *under a certain cancer treatment the overall liver toxicity always remains below a certain value*, and so on. In particular, we can make effective comparisons between alternative treatments.

We choose to model breast cancer, and consider different kinds of treatments/therapies, the size of cancer cells, and patient comorbidity. We explore how different variants of timed automata can be useful to capture different notions and allow us to make different inferences. All the variants discussed have tool support with advantages and limitations. We explore the possible combinations of the approaches which are more suited to provide answers to our questions. In this way we aim to find the best combination for a cancer automaton.

III. Our approach

Our approach models cancer progression using a mechanism to determine cancer growth based on the size of the primary tumour (T), the affected lymph nodes (N), and the presence of metastasis (M) called TNM Staging [7]. TNM staging has been widely used by physicians to plan the appropriate treatment for patients as well as estimating patient prognosis [8]. Our cancer automata are inspired by CHA but use TNM staging and are based on variants of timed automata [9] to obtain a (family of) cancer automata.

Timed automata (TA) [9] add the notion of time to standard automata (based on a finite set of states and labelled transitions between them) through a set of variables called *clocks*. Clocks are special variables which can be inspected or reset but not assigned a value. A time unit represents a second, minute or month, depending on what is a sensible unit for the model. A timing constraint can be placed on *locations* (the term used for states in a TA) to denote a *location invariant* (to indicate for instance how long the automaton can remain in the location) and on transitions where it acts as a guard. Although TA are very useful to model real-time systems, additional notions such as energy, memory consumption, band-width, cost, etc, cannot be captured directly. *Priced or weighted timed automata* (PTA) [10] have been introduced for this purpose and have additional special clock variables to capture time dependent quantities such as cost. In addition to cost, many real-life systems show stochastic behaviour (e.g., unreliable communication media due to bad weather, component failures, etc). This form of non-determinism is vital for many systems. *Probabilistic Timed Automata* (ProTA)[11] add probabilities to transitions. To also address real-time systems operating under uncontrollable environments, a further variant of timed automata was introduced called *Timed Game Automata* (TGA) [12]. A timed game can provide an algorithm that makes it possible to generate an adaptive strategy for both best and worst case scenarios caused by the system environment.

Cancer automata borrow the notions of *time constraint*, *clock*, and *location invariant* and use them to encode the properties to determine the progression of cancer. In general, an automaton consists of a finite set of *locations* which for a cancer automaton is restricted to the four locations associated to cancer TNM stages, where the last stage corresponds to metastasis. We also consider a further location to indicate when sufficient time has passed for the patient to be considered cancer free. Transitions are used to model the therapies and treatments for each stage, while location invariants are being used to determine the progression of cancer. Through verification of a cancer automaton (possibly in composition with additional automata for capturing other key aspects of a patient including any concomitant conditions), it is possible to select and evaluate amongst several treatments with respect to overall benefits (e.g., less impact on concomitant conditions) or other factors such as cost.

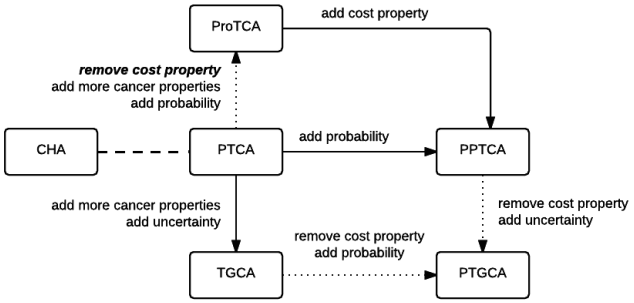


Fig. 1. Cancer automata framework

Our approach starts with Priced Timed Cancer Automata (PTCA) which is closest to the original timed CHA but with some modifications and contains a way of capturing cost (see Fig. 1). We then add more properties of cancer (e.g. tumour size, affected lymphatic nodes (NX)) as location invariants. We create a Probabilistic Timed Cancer Automata (ProTCA) by adding probabilities and removing the cost property. We combine both PTCA and ProTCA to obtain a Priced Probabilistic Timed Cancer Automata (PPTCA). Similarly, we add uncertainty into the PTCA model to obtain a Timed Game Cancer Automata (TGCA). Using TGCA, we are interested in finding a strategy which avoids metastasis without considering cost. Lastly, we obtain a new Probabilistic Timed Game Cancer Automata (PTGCA) by combining

both TGCA and PTGCA. With this model, we hope to decrease the uncertainty in accordance to the properties we have considered. The PTGCA has both branching transitions and uncontrollable transitions. In this model, a therapy may or may not improve the patient's condition. The likelihood of delaying or facilitating cancer progression is unknown.

IV. Cancer Analysis through Model Checking

We have used several variants of timed automata to explicitly explore different properties, concerns and associated tools when modelling (breast) cancer. These models allow us to separate different aspects which when combined result in a more expressive but more complex model. We reflect on the results of the analysis for PTCA, ProTCA, and PPTCA[2].

A. Priced Timed Cancer Automata (PTCA)

In this automaton, locations represent the stages of cancer, and transitions model cancer therapies. A therapy is represented by a transition that modifies the value of a global double variable $rate$. How the value of $rate$ changes depends on the therapy associated to the transition. If a medication is effective it delays the progression of cancer. Decreasing the (growth) $rate$ at a particular location (cancer stage) models a delay of the cancer growth. Each transition has a cost property that gradually increases and models the resources needed in order to perform the therapy. These cost properties contribute to the total cost value of the system. The increment depends on whether it has previously been applied to the patient as well as the progression of cancer (i.e. the higher the stage, the higher the cost property). In addition to locations that denote the stages of cancer, we have another location to represent a state where the treatment has been successful and it is possible to refer to a cancer free state (`CanFree`).

The goal of the PTCA model is to find the optimal cost to reach `CanFree` which can be specified in CTL by $\exists\Diamond$`CanFree` (there exists a path where eventually the location `CanFree` is reached). If existing, the optimal path is shown.

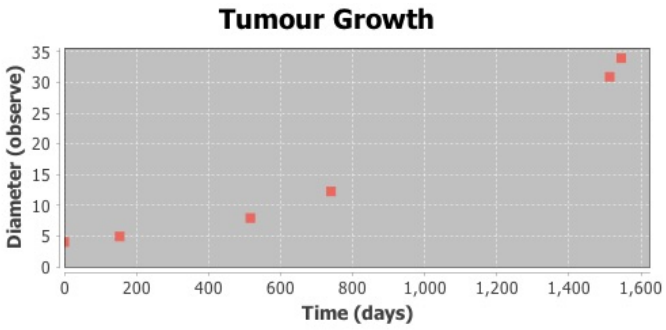


Fig. 2. Tumour Growth Case No. 1

B. Probability Timed Cancer Automata (ProTCA)

We extend our cancer model by considering more properties (e.g. tumour size, affected lymphatic nodes), which give us more accurate information on the stage of cancer. Our ProTCA models the cancer growth by using its tumour size and the spreading to nearby lymphatic nodes as the location invariants. To begin with, we have an integer variable `logD` to determine the size of the tumour in logarithmic scale. We use the tumour size of a patient data (i.e Case No. 1) in [13].

Fig. 2 plots the exponential growth of the tumour size. To simplify the modelling of the tumour growth, we scale the growth logarithmically. We perform a linear regression to get the polynomial equation for our ProTCA (i.e. $size = 0.041072500921616864 * time + 1.4017231839892936$). We then scale (i.e. by factor 100) and round the equation to implement it in our model. Then, we create an action `setGrowth()` in our ProTCA to simulate the growth. Our ProTCA model is too large to be shown here. The main idea is that since the success rate of a therapy is not exact [14], we use branching weight probabilities to express therapy effectiveness.

To take into account patient comorbidities [5], we can, for instance, model the levels of toxicity of the liver in a separate automaton. Liver toxicity is affected by external factors (e.g., additional medication) which may force transitions in the liver automaton. Conversely, if the liver has high toxicity, some therapies cannot be given because toxicity levels act as guards for some of the transitions in the cancer automaton.

C. Priced Probabilistic Timed Cancer Automata (PPTCA)

After we model the ProTCA, we reintroduce the cost property to the model. We assign a particular cost value to each therapy based on their effectiveness (as in the PTCA). The value denotes the resources needed for treating cancer (e.g. financial cost, method of treatment, medications).

The obtained PPTCA is an extension of the ProTCA with an additional cost property. We add further so-called strategy templates to model an appropriate set of actions to be performed in each stage of cancer. A strategy is a simple timed automaton which synchronises with the main cancer automaton through channels. A strategy can show alternatives if some therapies cannot be applied to the patient because of the comorbidity factor.

D. PPTCA Properties

To verify the PPTCA, we use logics TCTL and PCTL. We observe the estimation of the expected minimum and maximum value of the cost needed to perform a particular strategy. Each combination of strategies is monitored and compared. Next, we simulate the tumour growth, the affected lymphatic nodes (`NX`), the metastasis stage (`Stage4`), as well as the liver comorbidity. Some of our properties for the PPTCA are shown in Table I. Each query in Table I has been verified for all strategies. We are interested in comparing the cost and success rate of each strategy to avoid metastasis.

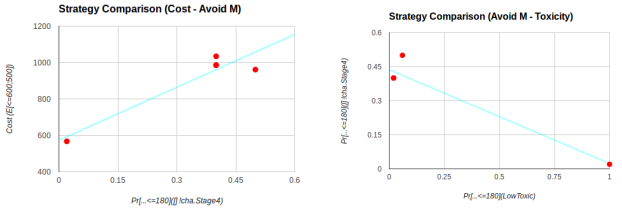
Through verification, we are able to compare some quantitative properties for each strategy. Further, we want to compare how optimal strategies are. The comparison is shown in both Fig. 3(a) and Fig. 3(b).

Fig. 3(a) shows the effectiveness of each strategy to avoid metastasis as well as the cost needed to perform

[2]See https://juliana.host.cs.st-andrews.ac.uk/CancerAutomata/ for details.

TABLE I
PPTCA PROPERTY SPECIFICATION

CTL/PCTL/Query Language	Remark
$EG\neg Stage4$	There exists a path where metastasis is never reached.
$P(F^{xrm\leq 300} Stage4)$	The probability of cancer cells reaching metastasis in less than 25 years.
$P(G^{xrm\leq 300} \neg Stage4)$	The probability of not reaching metastasis in less than 25 years.
$P(G^{xrm\leq 300} \neg LowToxic)$	The probability of always having low liver toxicity in less than 25 years.
$E[<= 600; 500](max : cha.cost)$	The estimation of the maximum cost within 600 time units in 500 runs.
`simulate 1 [<=200] (logD, lToxic*1000)`	Plotting tumour growth and liver toxicity condition. We multiply the boolean value lToxic by 1000 for easier observation.
`simulate 1 [<=200] (NX, Stage4*3)`	Plotting Metastasis and affected lymphatic nodes. We multiply `Stage4` by 3 for easier observation.



(a) Strategy Cost-Effectiveness (b) Strategy Effectiveness-Risk

Fig. 3. Strategy Comparison

each strategy. Although, Fig. 3(a) shows an upward trend in the probability of avoiding metastasis towards the cost (i.e. the higher the cost, the higher the probability of avoiding metastasis), this statement is not necessarily true. First of all, we need more data and strategies to correctly determine the relation between the cost property and the effectiveness of a strategy. However, adding more therapies may lead to the state-space explosion problem. To avoid the state-space explosion problem we can add only a subset of therapies used in the strategy to the PPTCA model (basically do the analysis in stages).

Fig. 3(b) shows the effectiveness of therapies against patient comorbidity (here only liver toxicity). From the strategy comparison shown in both Fig. 3(b) and Fig. 3(a), it is possible to choose a strategy based on patient preferences. We have also considered other cancer progression characteristics such as the spreading of cancer to the lymphatic nodes before metastasis, tumour growth, but we omit details here.

V. CONCLUSION

Our main goal was to explore the suitability of model checking as an automated technique to build a framework to quantify the suitability and effect of different treatment options for a specific patient with potential comorbidities. To have access to existing verification tools, we explored a new family of cancer automata obtained by extending variants of timed automata. Our models include different properties and information about cancer progression. By using branching weighted probabilities, for instance, we can model the prognosis rate [8] to some extent. Branching probabilities can be useful to model the effectiveness of a therapy towards the inhibition of cancer growth in comparison to a controlled transition. In addition, we introduced strategies, such as applying different therapies or medication based on cancer progression and comorbidity (i.e. liver toxicity), to give an idea of a treatment's effectiveness and possible risks.

Longitudinal data sets can be used to give real accurate values and probabilities to our model, making our framework more valuable and realistic. Breast cancer is one example where longitudinal data sets exist. Relevant work at this level is given in [15] where real data has been used to inform Markov-based models. This work shows how predicting the progression of metastatic cancer may contribute to chances of survival. Finally, we want to add more properties of cancer to our model and work closely with oncologists to understand the overall benefit of the approach.

REFERENCES

[1] National Cancer Institute: What Is Cancer. (2015) http://www.cancer.gov/cancertopics/what-is-cancer.

[2] McPherson, K., Steel, C.M., Dixon, J.M.: Breast cancer—epidemiology, risk factors, and genetics. BMJ **321** (2000) 624–628

[3] American Cancer Society: Cancer facts and figures (2015) Atlanta.

[4] Loohuis, L.O., Witzel, A., Mishra, B.: Cancer hybrid automata: Model, beliefs and therapy. Journal of Information and Computation **236** (2014) 68–86

[5] Sgaard, M., Thomsen, R.W., Bossen, K.S., Srensen, H.T., Nrgaard1, M.: The impact of comorbidity on cancer survival: a review. Clinical Epidemiology **5** (2013) 3–29

[6] Clarke, E.M., Emerson, E.A.: Design and synthesis of synchronization skeletons using branching time temporal logic. In: Logic of Programs. Volume 131 of Lecture Notes in Computer Science. Springer-Verlag (1981) 5271

[7] Greene, F., Balch, C., Haller, D., Morrow, M.: AJCC Cancer Staging Manual (6th Edition). Springer (2002)

[8] DeSantis, C., Lin, C., Mariotto, A., Siegel, R., Stein, K., Kramer, J., Alteri, R., Robbins, A., Jemal, A.: Cancer treatment and survivorship statistics, 2014. CA: A Cancer Journal for Clinicians **64** (2014) 252–271

[9] Alur, R., Dill, D.L.: A theory of timed automata. Theoretical Computer Science **126** (1994) 183–235

[10] Bouyer, P., Fahrenberg, U., Larsen, K.G., Markey, N.: Quantitative analysis of real-time systems using priced timed automata. Commun. ACM **54** (2011) 78–87

[11] Norman, G., Parker, D., Sproston, J.: Model checking for probabilistic timed automata. Formal Methods in System Design **43** (2013) 164–190

[12] de Alfaro, L., Faella, M., Henzinger, T., Majumdar, R., Stoelinga, M.: The element of surprise in timed games. In Amadio, R., Lugiez, D., eds.: CONCUR 2003 - Concurrency Theory. Volume 2761 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2003) 144–158

[13] Fournier, D.V., Weber, E., Hoeffken, W., Bauer, M., Kubli, F., Barth, V.: Growth rate of 147 mammary carcinomas. Cancer **45** (1980) 2198–2207

[14] Early Breast Cancer Trialists Collaborative Group: Effect of radiotherapy after breast conserving surgery on 10 years recurrence and 15 years breast cancer death. Breast Diseases: A Year Book Quarterly **23** (2012) 266–267

[15] Newton, P., Mason, J., Venkatappa, N., Jochelson, M., Hurt, B., Nieva, J., Comen, E., Norton, L., Kuhn, P.: Spatiotemporal progression of metastatic breast cancer: a markov chain model highlighting the role of early metastatic sites. npj Breast Cancer (2015)