

# An efficient algorithm for computing fixed length attractors based on bounded model checking in synchronous Boolean networks with biochemical applications

X.Y. Li<sup>1,2</sup>, G.W. Yang<sup>1</sup>, D.S. Zheng<sup>1,2</sup>, W.S. Guo<sup>1</sup> and W.N.N. Hung<sup>3</sup>

<sup>1</sup>School of Computer Science and Engineering,  
University of Electronic Science and Technology of China, Chengdu, China

<sup>2</sup>Department of Electronic Engineering,  
University of California at Los Angeles, Los Angeles, CA, USA

<sup>3</sup>Synopsys Inc., Mountain View, CA, USA

Corresponding authors: X.Y. Li / G.W. Yang

E-mail: erin.xiaoyu.li@gmail.com / guowu@uestc.edu.cn

Genet. Mol. Res. 14 (2): 4238-4244 (2015)

Received October 21, 2014

Accepted March 31, 2015

Published April 28, 2015

DOI <http://dx.doi.org/10.4238/2015.April.28.5>

**ABSTRACT.** Genetic regulatory networks are the key to understanding biochemical systems. One condition of the genetic regulatory network under different living environments can be modeled as a synchronous Boolean network. The attractors of these Boolean networks will help biologists to identify determinant and stable factors. Existing methods identify attractors based on a random initial state or the entire state simultaneously. They cannot identify the fixed length attractors directly. The complexity of including time increases exponentially with respect to the attractor number and length of attractors. This study used the bounded model checking to quickly locate fixed length attractors. Based on the SAT solver, we propose a new algorithm for efficiently computing the fixed length attractors, which is more suitable for large Boolean networks and numerous attractors' networks. After comparison

using the tool BooleNet, empirical experiments involving biochemical systems demonstrated the feasibility and efficiency of our approach.

**Key words:** Bounded model checking; Genetic regulatory network; Attractor; SAT solver; Synchronous Boolean network

## INTRODUCTION

Currently biologists use Boolean networks (Glass, 1985) to describe biological regulatory networks that function in biochemical systems to thoroughly understand the operating mechanisms of organisms. Kauffman (1995) previously described this phenomenon. In General, we can consider one condition of an organism and express its behavior using the genetic regulatory networks (GRNs). Next, the GRN can be modeled as a Boolean network. The genes follow Boolean rules to activate (1) and inhibit (0) the following genes according to the activities of their molecular inputs. All activities will follow a finite trajectory in their state space. Ultimately, the trajectory converges onto attractors that stabilize the system. Therefore, identifying the attractor is important for increasing the understanding of biochemical systems, particularly those in the cancer cells.

For the same GRNs, different modeling methods can be used, such as synchronous (Farrow et al., 2004; Fauré et al., 2006; Remy et al., 2006), asynchronous (Fauré et al., 2006; Garg et al., 2008) and semi-asynchronous (Fauré et al., 2006) models. Dubrova et al. (2005) presented a basic algorithm known as BooleNet to identify attractors in synchronous Boolean networks. Zhao (2005) demonstrated that the computing attractors in synchronous Boolean networks represent a nondeterministic polynomial time complete problem. Ay et al. (2009) developed a faster method for identifying attractors of self-loops and simple-loops. All of these tools are based on binary decision diagram (BDD) package (Lee, 1959).

This study used bounded model checking (BMC) (Clarke et al., 2001; Biere et al., 1999) to search for attractors in synchronous Boolean networks. Combined with the BMC theory, we propose a new algorithm for quickly identifying fixed length attractors in synchronous Boolean networks based on the SAT solver (Moskewicz et al., 2001; Goldberg and Novikov, 2002). After comparison with BooleNet (Dubrova et al., 2005), our experimental results showed that our algorithm was more efficient and feasible for large synchronous Boolean networks; it can also be used for analyzing actual large-scale biochemical system models.

This paper is organized as follows: we first provide a brief review of the synchronous Boolean network and its basic definitions. In the following subsection, our new theory and an algorithm are explained. Next, a biological regulatory network is used as a simple example to demonstrate the computation of attractors. In the Results section, we use 7 benchmarks to demonstrate that our approach is highly efficient and feasible. The top 5 benchmarks are from realistic biochemical systems. The remaining are randomly generated benchmarks.

## MATERIAL AND METHODS

### Synchronous Boolean networks and basic definitions

In general, a synchronous Boolean network contains  $n$  nodes  $(x_1, x_2, x_3, \dots, x_n)$ , with each node updating itself with each other in a synchronous manner (Milligan and Wilson,

1993; Farrow et al., 2004). Each node has only 1 Boolean value: 1 (ON) or 0 (OFF) at 1 moment. Commonly, a synchronous Boolean network's value 1 (Active) or 0 (Inhibit) represents the status of its biological regulatory network. Therefore, we can describe the synchronous Boolean network using the following equations:

$$\begin{aligned}x_{1,t+1} &= f_1(x_{1,t}, x_{2,t}, x_{3,t}, \dots, x_{n,t}); \\x_{2,t+1} &= f_2(x_{1,t}, x_{2,t}, x_{3,t}, \dots, x_{n,t}); \\x_{3,t+1} &= f_3(x_{1,t}, x_{2,t}, x_{3,t}, \dots, x_{n,t}); \\&\dots \\x_{n,t+1} &= f_n(x_{1,t}, x_{2,t}, x_{3,t}, \dots, x_{n,t});\end{aligned}$$

The above equations can also be simplified as shown below:

$$X_{t+1} = F(X_t) \quad (\text{Equation 1})$$

Here,  $x_t = (x_{1,t}, x_{2,t}, x_{3,t}, \dots, x_{n,t})$  represents a state at time  $t$ ,  $x_{i,t} \in \{0,1\}$ . The translation function  $F = (f_1, f_2, f_3, \dots, f_n)$  is the Boolean function from  $\{0,1\}^n$  to  $\{0,1\}^n$ , that is, the state at time  $t + 1$  can be determined using the previous state at time  $t$ .

**Definition 1. Predecessor/Successor:** For a synchronous Boolean network with  $n$  nodes and a translation function  $F$ , if  $X_{t+1} = F(X_t)$ , state  $X_{t+1}$  is the successor of state  $X_t$  and state  $X_t$  is the predecessor of state  $X_{t+1}$ , respectively. Next, we use  $(X_t, X_{t+1})$  to represent that  $X_{t+1}$  is the successor state of  $X_t$ .

**Definition 2. An Attractor  $A_{it}$ :** a set of states to which each state can return itself after finite translation computing by  $F$ .  $Length(A_{it})$  is the state number of attractor  $A_{it}$ .  $U_{A_{it}}$  represents all attractors whose lengths are  $n$  in a Boolean network.

## Theoretical results

This subsection presents our theory and algorithm. We also demonstrate, using a biological regulatory network, how to compute the fixed length attractors. Equation 1 is the simplified form of the synchronous Boolean network. However, we can transform this equation to the following expression:

$$F(X_t, X_{t+1}) = 1 \quad (\text{Equation 2})$$

Here, all variables are the same as those in Equation 1. Based on the definition of an attractor (Definition 2), an attractor is a loop with finite states. In this loop, each state can return itself to a particular state through fixed transitions; that is, all states can return themselves after  $Length(A_{it})$  transitions. Therefore, we use the following equation to indicate an attractor:

$$T = \bigwedge_{i=m}^{m+length-1} F(X_i, X_{i+1}) \wedge (X_m \odot X_{m+length}) \quad (\text{Equation 3})$$

Equation 3,  $T$  is the propositional formula, which is unfolded from time step  $m$  to time step  $m+length-1$ , where  $m$  is a state at any time and  $length$  is an attractor length. Additionally, “ $\wedge$ ” and “ $\odot$ ” represent “AND” and “NOR” operators in Boolean logic, respectively. The propositional formula  $T$  represents an attractor, in which a state  $X_i$  will return to itself after length transition steps.

### Quickly locate fixed length attractors (QLFLA) algorithm

Based on BMC theory and Equations 3, we propose an algorithm for quickly locating fixed length attractors using the pseudo code shown below. In Algorithm 1, variable “ $A_u$ ” indicates any attractor whose length is  $n$ ; “ $A_{u\_num}$ ” is the number of attractors whose length is  $n$ ; “ $U_{Att}$ ” is the union set of attractors whose length is  $n$ . In the pseudo code, “ $Sat(T)$ ” corresponds to the so-called SAT solver. Furthermore, if there is any assignment of all variables, the SAT solver takes an expression and returns the result of true. Otherwise, it will return the result of false. In contrast, there is no assignment that qualifies this expression. The symbol “ $\leftarrow$ ” is assignment from left to right.

#### Algorithm 1 QLFLA on BMC

```

Input: Propositional formula  $T$  and length  $n$ .
01: Initial:
 $U_{Att} \leftarrow$  ;
 $A_u \leftarrow$  ;
 $A_{u\_num} \leftarrow 0$ ;
02: While ( $Sat(T)$ ) do
03:  $X_i \leftarrow Sat(T)$ ;
04:  $A_u \leftarrow A_u X_i$ ;
05: For ( $j \leftarrow 1; j < n; j \leftarrow j+1$ )
06:  $X_{j+1} \leftarrow F(X_j)$ ;
07:  $X_i \leftarrow X_{j+1}$ ;
08:  $A_u \leftarrow A_u X_i$ ;
09: End For
10:  $print(A_u)$ ;
11:  $U_{Att} \leftarrow U_{Att} A_u$ ;
12:  $A_{u\_num} \leftarrow A_{u\_num} + 1$ ;
13:  $T \leftarrow T \wedge \neg A_u$ ;
14:  $A_u \leftarrow$  ;
15: End While
16: Output: return  $A_{u\_num}, U_{Att}$ ;

```

### A biological regulatory network example

Based on previous studies (Glass and Kaufmann, 1973; Heidel et al., 2003), one condition of sigmoidal kinetics systems was chosen from a complicated environment, which is shown using the following differential equations:

$$\frac{dx_1}{dt} = \frac{x_2^2}{1+x_2^2} \cdot \frac{x_3^2}{1+x_3^2} - x_1;$$

$$\frac{dx_2}{dt} = \frac{x_1^2}{1+x_1^2} - x_2;$$

$$\frac{dx_3}{dt} = \frac{1}{1+x_2^2} - x_3;$$

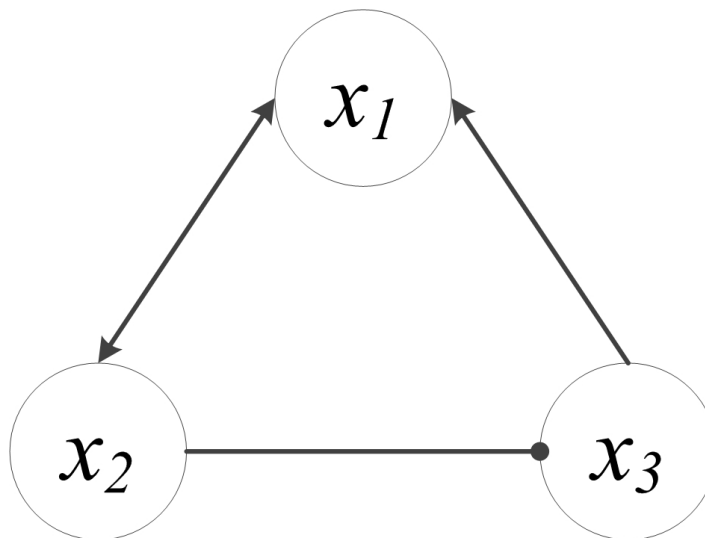
Based on a study of Heidel et al. (2003), we simplified the above differential equations as the Boolean network shown in Figure 1 (Boolean network) and its Boolean translation function. This is a nonlinear logic function with negative feedback.

$$x_{1,t+1} = x_{2,t} \wedge x_{3,t};$$

$$x_{2,t+1} = x_{1,t};$$

$$x_{3,t+1} = \neg x_{2,t};$$

In this example, there are 3 Boolean nodes:  $x_1$ ,  $x_2$ , and  $x_3$  and. If we wish to compute whether an attractor exists with length 2, the propositional formula can be used:  $T = (X_1 X_{i+1}) \wedge (X_{i+1} X_{i+2}) \wedge (X_i \odot X_{i+2})$ . After computing this propositional formula based on the SAT solver (Moskewicz et al., 2001; Goldberg and Novikov, 2002), we can easily obtain the attractor with the above conditions:  $(011,100) \wedge (100,011) \wedge (011 \odot 011)$ . Next, we obtain an attractor whose length is 2, as:  $(011) \rightarrow (100) \rightarrow (011)$ .



**Figure 1.** A Boolean network of the sigmoidal kinetics system. Node  $x_1$  activates node  $x_2$ ; node  $x_2$  inhibits  $x_3$ ; node  $x_3$  and  $x_2$  active node  $x_1$ . All nodes update themselves by synchronous mechanism.

## RESULTS

In this section, we compare our results and those of previous studies obtained using our algorithm (QLFLA) and BooleNet (Dubrova et al., 2005). The experimental results show that our algorithm is efficient and highly feasible particularly for large biological regulatory networks. The benchmarks contain 7 different test cases. The top 5 are Boolean models generated from realistic GRNs. The last 2 are randomly generated networks. All experiments can be performed on an Intel Core™ CPU 4300 1.80 GHz with 2 GB memory on an Ubuntu 9.04 Linux server.

The top 4 Boolean networks were generated from 5 organisms, including *Arabidopsis thaliana* (Chaos et al., 2006), budding yeast (Li et al., 2004), *Drosophila melanogaster* (Albert and Othmer, 2003), T-helper cell (Mendoza and Xenarios, 2006), and protein (Gonze and Goldbeter, 2001; Heidel et al., 2003). In addition to these 5 classical benchmarks, we provide the 2 other randomly generated Boolean networks. All experimental results are shown in Table 1.

**Table 1.** Experimental results of seven models.

Benchmark	Nodes	Time (s)		SCAs number		Length
		BooleNet	FSSBN	BooleNet	FSSBN	
		(Dubrova et al., 2005)		(Dubrova et al., 2005)		
<i>Arabidopsis thaliana</i> (Chaos et al., 2006)	15	0.024	0.011	10	10	1
Budding yeast (Li et al., 2004)	12	0.042	0.020	7	7	1
<i>Drosophila melanogaster</i> (Albert and Othmer, 2003)	52	Timeout	0.078	E	7	1
T-helper cell (Mendoza and Xenarios, 2006)	23	0.045	0.018	3	3	1
Protein (Gonze and Goldbeter, 2001)	30	Timeout	85.62	E	35790267	30
Random Network 1	1375	Timeout	0.103	E	0	1
Random Network 2	2095	Timeout	0.318	E	1	1

Table 1 shows the runtime of Algorithm 1 (QLFLA) and BooleNet (Dubrova et al., 2005) under the same conditions. “Timeout” represents the running time over  $12 \times 3600 = 43,200$  s. “E” indicates that no result was returned. The 4 testcases (*D. melanogaster*, protein, and random network 1 and 2) show that our algorithm can handle larger networks and compute their attractors in a short amount of time. However, BooleNet cannot be used to obtain these results. For the rest testcases, our algorithm is at least twice as faster as BooleNet. The experimental results demonstrate that our algorithm is highly efficient for the large Boolean network and for identifying numerous attractors in Boolean networks.

## DISCUSSION

We applied BMC to biological regulatory networks of biochemical systems. We also proposed a new algorithm for quickly locating fixed length attractors based on the SAT solver. In addition, we used 5 classical realistic biological networks and 5 random networks to demonstrate that our algorithm is efficient and highly feasible for examining larger networks and numerous attractors in the same Boolean networks. In the future, we will cooperate with biologist to apply our algorithm to larger and more realistic biological regulatory networks.

## ACKNOWLEDGMENTS

Research supported by the Fundamental Research Funds for the Central Universi-

ties (Grant #ZYGX2011YB022), the National Natural Science Foundation of China (Grant #61272175, and #973 Foundation #2010CB328004).

## REFERENCES

- Albert R and Othmer HG (2003). The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *J. Theor. Biol.* 223: 1-18.
- Ay F, Xu F and Kahveci T (2009). Scalable steady state analysis of Boolean biological regulatory networks. *PLoS One* 4: e7992.
- Biere A, Cimatti A, Clarke E and Zhu Y (1999). Symbolic Model Checking Without BDDs. In: Tools and Algorithms for the Construction and Analysis of Systems Lecture Notes in Computer Science (Cleaveland WR, ed.). Springer, New York, 193-207.
- Chaos A, Aldana M, Espinosa-Soto C and León BGP (2006). From genes to flower patterns and evolution: dynamic models of gene regulatory networks. *J. Plant Growth Regul.* 25: 278-289.
- Clarke E, Biere A, Raimi R and Zhu Y (2001). Bounded model checking using satisfiability solving. *Formal Methods System Design* 19: 7-34.
- Dubrova E, Teslenko M and Martinelli A (2005). Kauffman Networks: Analysis and Applications. In: 484 Proceedings of the 2005 IEEE/ACM International Conference on Computer-aided Design, IEEE Computer Society, Washington, 479.
- Farrow C, Heidel J, Maloney J and Rogers J (2004). Scalar equations for synchronous Boolean networks with biological applications. *IEEE Trans. Neural Netw.* 15: 348-354.
- Fauré A, Naldi A, Chaouiya C and Thieffry D (2006). Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics* 22: e124-e131.
- Garg A, Di Cara A, Xenarios I, Mendoza L, et al. (2008). Synchronous versus asynchronous modeling of gene regulatory networks. *Bioinformatics* 24: 1917-1925.
- Glass L (1985). Boolean and Continuous Models for the Generation of Biological Rhythms. In: Dynamical Systems and Cellular Automata (Demongeot J, Goles E and Tchuente M, eds.). Academic Press, London, 197-206.
- Glass L and Kauffman SA (1973). The logical analysis of continuous, non-linear biochemical control networks. *J. Theor. Biol.* 39: 103-129.
- Goldberg E and Novikov Y (2002). BerkMin: a Fast and Robust SAT-solver. Proceedings of Design, Automation and Test in Europe Conference, Washington, 142-149.
- Gonze D and Goldbeter A (2001). A model for a network of phosphorylation-dephosphorylation cycles displaying the dynamics of dominoes and clocks. *J. Theor. Biol.* 210: 167-186.
- Heidel J, Maloney J, Farrow C and Rogers J (2003). Finding cycles in synchronous Boolean networks with applications to biochemical systems. *Int. J. Bifurcat. Chaos* 13: 535-552.
- Kauffman S (1995). At Home in the Universe. Oxford University Press, New York.
- Lee C (1959). Representation of switching circuits by binary-decision programs. *Bell System Tech. J.* 38: 985-999.
- Li F, Long T, Lu Y, Ouyang Q, et al. (2004). The yeast cell-cycle network is robustly designed. *Proc. Natl. Acad. Sci. U. S. A.* 101: 4781-4786.
- Mendoza L and Xenarios I (2006). A method for the generation of standardized qualitative dynamical systems of regulatory networks. *Theor. Biol. Med. Model.* 3: 13.
- Milligan DK and Wilson MJD (1993). The behaviour of affine boolean sequential networks. *Conn. Sci.* 5: 153-167.
- Moskewicz MW, Madigan CF, Zhao Y and Zhang L (2001). Chaff: Engineering an Efficient SAT Solver. Proceedings of the 38th Design Automation Conference, 530-535.
- Remy E, Ruet P, Mendoza L and Thieffry D (2006). From logical regulatory graphs to standard petri nets: dynamical roles and functionality of feedback circuits. *Trans. Comput. Syst. Biol.* 4320: 56-72.
- Zhao Q (2005). A remark on "scalar equations for synchronous Boolean networks with biological applications" by C. Farrow, J. Heidel, J. Maloney, and J. Rogers. *IEEE Trans. Neural Netw.* 16: 1715-1716.