Implicit Enumeration of Strongly Connected Components [†]

Aiguo Xie[‡] and Peter A. Beerel
Department of Electrical Engineering - Systems
University of Southern California
Los Angeles, CA 90089-2562

Abstract

This paper presents a BDD-based implicit algorithm to compute all maximal strongly connected components of directed graphs. The algorithm iteratively applies reachability analysis and sequentially identifies SCCs. Experiments suggest that the algorithm dramatically outperforms the only existing implicit method which must compute the transitive closure of the adjacency-matrix of the graphs.

1 Introduction

Decomposing a directed graph (digraphs) into its (maximal) strongly connected components (SCCs) is a fundamental graph problem [1] and has many important applications in CAD. Generally speaking, SCC decomposition often divides a digraph problem into subproblems, one for each SCC. The solution to the original problem can be constructed by combining the solutions to the subproblems, sometimes with the aid of the component graph (i.e., the structure of connections among SCCs).

Using an explicit data structure such as an *adjacency-list* or an *adjacency-matrix* [1], the decomposition of a digraph $G(V, E)$ (with V being the set of its nodes and E the set of its edges) can be done in linear time (i.e., $O(|V| + |E|)$) using a depth-first search [1, 2]. However, in many real application, the size of the digraph can be too large (e.g. with more than 10^{20} nodes) for explicit methods to be practical.

To handle larger problems, an implicit method using BDDs [3] is described in [4] to find all SCCs in the reachable state space of a finite state machine[1]. It first computes the transitive closure of the state transition relation of the machine, and then computes all SCCs *simultaneously*[2]. We call this method TC-based. In practice, however, computing the transitive closure has been shown to be very computationally expensive in both CPU time and memory.

This paper proposes an alternative approach that is motivated by related algorithms that identify only the *terminal* SCCs, i.e., those from which no other SCCs can be reached [6, 7, 8]. These algorithms, rather than extracting the terminal SCCs from the set of all SCCs found using the above TC-based method, identify all terminal SCCs *sequentially* by iteratively applying reachability analysis [6, 7, 8]. Because reachability analysis is much less computationally expensive than computing the transitive closure and finite state machines often have a limited number of terminal SCCs, these algorithms have been shown to be much more efficient than the TC-based method in finding terminal SCCs.

In particular, this paper explores whether such a BDD-based reachability analysis approach can be extended to sequentially find *all* SCCs (not just the terminal SCCs) of a digraph. This paper answers this question in the affirmative by presenting an algorithm that recursively partitions the digraph into subgraphs using reachability analysis and reduces the SCC identification problem to individual subgraphs that are sequentially analyzed. Experiments show that the algorithm is dramatically faster and solves much larger problems than the TC-based method, especially when the graph has a small number of SCCs.

The remainder of this paper is organized as follows. Section 2 reviews implicit representation of digraphs using BDDs, and summaries the transitive-closure-based method (TC-based method below) for computing all SCCs. Section 3 details our reachability-analysis-based method (RA-based method below). The performance of the two methods are compared in Section 4. We conclude the paper in Section 5, discussing potential applications our algorithm to system analysis and formal verification.

2 Preliminaries and the TC-based method

Let $G(V, E)$ denote a digraph where V is the set of its nodes, and for any two nodes $u, v \in V$, $(u, v) \in E$ iff there is an edge from u to v. A vector $(v_1, v_2, \cdots, v_{n+1})$ of nodes is a *path* of length $n+1$ (from v_1 to v_{n+1}) iff $(v_i, v_{i+1}) \in E$, for all $i = 1, \cdots, n$. Node v is *reachable* from u (denoted by $u \rightsquigarrow v$) if there is a path from u to v. A (maximal) *strongly connected component* (SCC) is a (maximal) subset of nodes where every pair of nodes are reachable from each other. Below, when we say SCCs, we refer to the maximal ones. We denote by $\mathcal{A}(G)$ the set of SCCs in G. For convenience, we call a node a non-SCC node if it does not belong to any SCC, an SCC node otherwise.

Let the nodes be labeled with distinct numbers from $\{1, 2, \cdots, |V|\}$. The digraph can then be represented by an adjacency matrix $M_{|V| \times |V|}$ [1] whose element $M(u, v)$ is 1 if $(u, v) \in E$, and 0 otherwise. The digraph may also be represented by a BDD $N(X, Y)$ by encoding the rows and columns of its adjacency matrix M with two sets of BDD variables X and Y. The BDD $N(X(u), Y(v))$ evaluates to 1 if $M(u, v) = 1$ where $X(u)$ and $Y(v)$ are the vectors encoding the labels of u and v, respectively. Details of BDDs and their common operations can be found in [3]. For convenience, we sometimes use node u to refer to the BDD encoding of its labeling.

The TC-based symbolic method first computes the transitive closure of N, denoted by N^*. By definition, $N^*(u, v) = 1$ iff $u \rightsquigarrow v$. The transpose of N^* denoted by N^{*t} has the property that $N^{*t}(u, v) = 1$ iff $v \rightsquigarrow u$. Thus, u and v belong

[†]This work is funded in part by a NSF CAREER Award MIP-9502386 and an SRC grant #98-DJ-486.

[‡]On leave for Cadence Design Systems, Inc., San Jose, CA.

[1]The method serves as an intermediate step in finding a subclass of SCCs called terminal SCCs in the paper.

[2]A similar idea is also exploited in [5] to find terminal SCCs of states in a finite state machine.

to a same SCC iff $N^*(u,v) \wedge N^{*t}(u,v) = 1$, where $\wedge$ denotes the BDD AND [3]. Consequently, the union of all SCC states (denoted by H) can be computed as $H = \exists_Y N^* \wedge N^{*t}(X,Y)$, where $\exists$ denotes the BDD *existential quantification* [3]. Finally, the SCC containing a particular state v in H can be computed as $\exists_Y X(v) \wedge N^* \wedge N^{*t}(X,Y)$.

3 The RA-based method

Computing the transitive closure of the adjacency-matrix of digraph as required by the TC-based method is equivalent to computing the reachability set of every node (the set of nodes reachable from the given node). Our method which is based on reachability analysis (the RA-based method below) needs to compute the reachability sets of potentially very few nodes. To better describe our method, we need to introduce a few notations related to the reachability sets and several properties of such sets.

3.1 Forward and backward sets

The forward set $F(v)$ of a node $v \in V$ is the set of nodes reachable from v. That is, $F(v) = \{u \in V \mid v \rightsquigarrow u\}$. Similarly, its backward set $B(v)$ is the set of nodes that can reach v. That is, $B(v) = \{u \in V \mid u \rightsquigarrow v\}$. One of the properties of $F(v)$ and $B(v)$ is given in Lemma 3.1, which states that the intersection of the two sets (if not empty) is an SCC. Additionally, v is a non-SCC node if the intersection is empty.

Lemma 3.1 *[6, 8, 7, 5] If v is a node of $G(V,E)$, then $F(v) \cap B(v) \in \mathcal{A}(G)$.*

A subset $U \subseteq V$ is *SCC-closed* if no SCC intersects with both U and its complement $V \backslash U$. That is, any SCC must be either completely contained in such a set or they do not overlap.

Lemma 3.2 *Both backward and forward sets of any node are SCC-closed.*

Proof: Since a digraph and its (edge-) reversed graph have the same set of SCCs, we just need to show that any backward set is SCC-closed. Consider a node u, and assume its backward set $B(u)$ is not SCC-closed. Thus, there is a non-empty SCC, say A such that $A \cap B(u) \neq \emptyset$ and $A \backslash B(u) \neq \emptyset$. Let x and y be any nodes of $A \cap B(u)$ and $A \backslash B(u)$, respectively. Since x and y are nodes of A, y must be a node of $B(x)$. Moreover, since x is a node of $B(u)$, $B(x) \subseteq B(u)$ [6]. Thus, y is a node of $B(u)$, which means $A \backslash B(u) \subseteq B(u)$. The latter can be true only if $A \subseteq B(u)$ or $A = \emptyset$, which contradicts the assumption. □

Theorem 3.1 *The difference of the backward sets of any two nodes is SCC-closed.*

Proof: Let u and v be any two nodes. Without loss of generality, let us prove the difference $D = B(v) \backslash B(u)$ is SCC-closed. If $D = B(v)$ or $D = \emptyset$, the result is trivial. Now suppose $D \neq B(v)$ and $D \neq \emptyset$, but it is not SCC-closed. Then, there must be an SCC, say A such that $A \cap D \neq \emptyset$ and $A \backslash D \neq \emptyset$. Since $B(v)$ is SCC-closed by Lemma 3.2, we have $A \subseteq B(v)$, and thus $A \backslash D \subseteq B(v)$. Consequently, $\emptyset \neq A \backslash D \subseteq B(v) \backslash D \subseteq B(u)$. Therefore, $A \cap B(u) \neq \emptyset$. Besides, $\emptyset \neq A \cap D = A \cap (B(v) \backslash B(u)) \subseteq A \cap (V \backslash B(u)) = A \backslash B(u)$. Hence, $B(u)$ is not SCC-closed. This contradicts Lemma 3.2. □

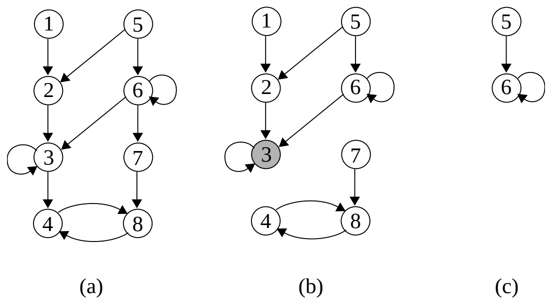

Figure 1: A digraph and its decomposition.

Remark: By similar arguments, one can show that the difference of any two of $B(u)$, $B(v)$, $F(u)$ and $F(v)$ is SCC-closed, and more generally, that the difference of any two SCC-closed sets is SCC-closed.
Example: In the digraph shown in Figure 1(a), $B(3) = \{1,2,3,5,6\}$, $F(3) = \{3,4,8\}$, and $B(4) = \{1,2,3,4,5,6,7,8\}$. From Lemma 3.1, $B(3) \cap F(3) = \{3\}$ is an SCC. From Lemma 3.2, $B(3)$ is SCC-closed. Moreover, the difference $B(4) \backslash B(3) = \{4,7,8\}$ is also SCC-closed from Theorem 3.1.

3.2 The algorithm

Lemma 3.2 and Theorem 3.1 suggest that a digraph may be first partitioned into backward sets (or differences of backward sets) of some of its nodes, and then the computing of SCCs of the graph can be restricted to each of the partitions. This leads to a divide-and-conquer strategy. For instance, the digraph in Figure 1(a) can be partitioned into $B(3)$ and $B(4) \backslash B(3)$, which reduces the problem to finding SCCs within the two subsets independently, as illustrated in Figure 1(b).

Because a digraph G and its transpose G^t (the graph obtained by reversing all the edges of G) has the same set of SCCs [1], the above divide-and-conquer strategy may also be carried out in term of forward sets. In fact, the partitioning is also possible to be carried out by interleaving the computation of backward sets and forward sets. However, our experience shows that none of these three different partitioning schemes leads to a significant advantage over the others. For this reason, we focus on explaining the algorithm based on the backward-set partitioning scheme. We note that the above divide-and-conquer strategy may also be implemented using an explicit data structure, but would be impractical for large graphs as in the explicit depth-first-search algorithm.

At the top level, the algorithm first picks a node from the set of remaining nodes (the entire set V at the beginning), computes its backward set restricted to the remaining nodes, and then decomposes the backward set into SCCs and a set of non-SCC nodes. This process repeats until the remaining set of nodes is empty.

The procedure is formally described in Figure 2. In the algorithm, function `random_take`(V') randomly picks a node from the set V'. Function `backward_set`(v, V', N) returns the backward set of v *restricted to* set V', which can be implemented efficiently using the fixed-point computation [9]. The remainder of this subsection explains the core procedure `SCC_DECOMP_RECUR`(v, $B(v)$, N) which recursively decomposes $B(v)$ into SCCs and a set of non-SCC nodes.

To describe the procedure `SCC_DECOMP_RECUR`, we introduce a concept of predecessors with finite maximum distance, also referred to as *FMD predecessors*. A node u is

```
SCC_DECOMP($N, V$)
    $V' \leftarrow V$;
    while($V' \neq 0$) {
        $v \leftarrow$ random_take($V'$);
        $B(v) \leftarrow$ backward_set($v, V', N$);
        SCC_DECOMP_RECUR($v, B(v), N$);
        $V' \leftarrow V' \wedge \overline{v \vee B(v)}$;
    }
```

Figure 2: The top-level algorithm.

```
FMD_PRED($W, U, N$)
    $pred \leftarrow 0$;
    $front \leftarrow W$;
    $bound \leftarrow U$;
    while($front \neq 0$) {
        $x \leftarrow \exists_Y front(Y) \wedge N(X,Y) \wedge bound(X)$;
        $y \leftarrow \exists_Y bound(Y) \wedge N(X,Y) \wedge bound(X)$;
        $front \leftarrow x \wedge \overline{y}$;
        $pred \leftarrow pred \vee front$;
        $bound \leftarrow bound \wedge \overline{front}$;
    }
    return $pred$;
```

Figure 3: Computing the predecessors with finite maximum distance.

an FMD predecessor of node v if any path from u to v has finite length. That is to say, any path from u to v must not pass any SCC. A necessary condition for this is that u must be a non-SCC node. The set of FMD predecessors of a set W of nodes is defined to be the set of all FMD predecessors of nodes in W, denoted by $FMD\_PRED(W)$. As an example, in Figure 1(b), $FMD\_PRED(\{3\}) = \{1, 2\}$. Node 5 is not a FMD predecessors of node 3 because it may pass the SCC $\{6\}$ to reach node 3. Figure 3 gives an implicit algorithm that iteratively computes the set $FMD\_PRED(W)$ restricted to a set U.

The recursive procedure SCC_DECOMP_RECUR(v, $B(v)$, N) works as follows. It first computes the forward set $F(v)$ of v restricted to $B(v)$. If $F(v)$ is not empty, then it is an SCC due to Lemma 3.1. Otherwise, node v is a non-SCC node. In either case, $F(v) \vee v$ can be removed from $B(v)$ from further consideration. Next, the FMD predecessors of $F(v) \vee v$ is computed, and are subsequently removed from further consideration. Then, from the remaining set of nodes, the procedure computes the set of immediate predecessors (IP) of these already removed nodes. To decompose the remaining set of nodes, a node u is randomly picked from the IP set and its backward set $B(u)$ is computed. The procedure then calls itself to decompose $B(u)$. After it returns, u and $B(u)$ are removed from the remaining set of nodes and the IP set gets updated. If the remaining set of nodes is not empty, another node w is picked up from the updated IP set, and its backward set $B(w)$ is recursively decomposed. This process repeats until the remaining set of nodes is empty. A formal description of the procedure is given in Figure 4.

Example: In figure 1(b), a call to SCC_DECOMP_RECUR(3, B(3), N) first computes $F(3, B(3), N) = \{3\}$, and reports $F(3) \cap B(3) = \{3\}$ to be an SCC. Next, it computes $FMD\_PRED(\{3\}) = \{1, 2\}$ which is reported to be a set of non-SCC nodes, and the IP set is computed to be $\{5, 6\}$. At this point, the procedure has reduced the problem to decomposing the digraph in Figure 1(c). Suppose node 6 is picked from the IP set, the procedure calls

```
SCC_DECOMP_RECUR($v, B(v), N$)
    $F(v) \leftarrow$ forward_set($v, B(v), N$);
    if($F(v) \neq 0$)
        report $F(v)$ an SCC;
    else
        report $v$ non-SCC;
    $x \leftarrow F(v) \vee v$;
    $R \leftarrow B(v) \wedge \overline{x}$;
    $y \leftarrow$ FMD_PRED($x, R, N$);
    report $y$ non-SCC;
    $R \leftarrow R \wedge \overline{y}$;
    $IP \leftarrow \exists_Y (y \vee x)(Y) \wedge N(X,Y) \wedge R(X)$;
    while($R \neq 0$) {
        $v \leftarrow$ random_take($IP$);
        $B(v) \leftarrow$ backward_set($v, R, N$));
        SCC_DECOMP_RECUR($v, B(v)$));
        $R \leftarrow R \wedge \overline{v \vee B(v)}$;
        $IP \leftarrow IP \wedge \overline{v \vee B(v)}$;
    }
```

Figure 4: Recursive decomposition of a backward set.

SCC_DECOMP_RECUR(6,$B(6)$,N) which, in this case, leaves no more nodes to be decomposed when it returns.

3.3 The complexity

As shown in Figure 4, procedure SCC_DECOMP_RECUR performs a constant number of reachability analyses before it calls itself. Each reachability analysis must have completed in $O(\mathcal{D})$ BDD operations where $\mathcal{D}$ is the diameter of the graph. Since a call to SCC_DECOMP_RECUR removes at least one node from further consideration, SCC_DECOMP_RECUR(v,$B(v)$,N) cannot call itself for more than $|B(v)|$ times. Thus, a trivial upper bound on the complexity of our algorithm is $O(|V| \times \mathcal{D})$ in BDD operations.

From our experiments, we observe that a call to SCC_DECOMP_RECUR typically removes an SCC before the procedure calls itself. In practice, the algorithm terminates in $O(|\mathcal{A}(G)| \times \mathcal{D})$ time (in BDD operations). Note further that the reachability analysis performed by our algorithm is always restricted to the subset of the nodes that is currently under consideration. In other words, the analysis needs a potentially much smaller number of BDDs operations than the diameter of the graph. In this respect, our above complexity analysis is pessimistic.

4 Experiments

We have implemented both TC-based and RA-based methods with `vis-1.3` [10]. In particular, the transitive closure of adjacency-matrix of the graph is computed by iteratively squaring the matrix. This section compares their performance using the examples distributed with the `vis-1.3` package each of which specifies a finite state machine. The methods are used to compute the SCCs in the reachable state space of the finite state machines. The package is run on a Sun Ultra 10 with 640 Mbytes of memory. Table 1 gives the experimental results of a representative set of the examples.

Since the TC_method simultaneously computes all SCCs, we report its run time as soon as it computes the transition closure of the adjacency-matrix and ANDs the transitive closure with its transpose. That is, the time to list all the SCCs is not included for in the reported run times of the

Example	State space			CPU secs	
	$\|V\|$	$\|V'\|$	SCCs	TC	RA
abp	484	444	11	5.41	0.40
arbit	5,568	5,568	1	797	0.52
bakery	2,886	2,604	31	197	3.24
coherence	94,748	94,688	2	`mo`	41.0
cdnew	186,876	139,520	3,469	`to`	202
eisenberg	1,611	1,609	1	`to`	0.49
emodel	34,133	376	12	to	96.0
scheduler	2.4e+6	2.4e+6	1	`mo`	1.19
minMax8	2.8e+6	2.8e+6	1	`mo`	0.37
minMax16	4.7e+13	4.7e+13	1	`mo`	0.84
minMax32	1.3e+28	1.3e+28	1	`mo`	8.50
tcp	3.9e+22	-	-	`mo`	`to`

Table 1: The experimental results where $|V|$ is the number of reachable states, $|V'|$ denotes the number of states belonging to SCCs, `mo` denotes memory out and `to` denotes time out after 1 hour of CPU time.

TC_method.

As shown in the Table, the TC-method solves only a few small examples whereas our RA-based method solves most of them. Even for those solved by the TC-based method, the RA-based method is faster by orders of magnitude. The RA-based method runs out of time in the last example. The reason is that the example contains excessive SCCs and the RA-based method cannot finish reporting all of the them in time, which highlights the limitation of our method.

Note also that for some of the above examples that have relatively small number of nodes, explicit methods might be more efficient than our RA-based method. However, for any large example, explicit methods are simply not applicable.

5 Conclusion

Computing strongly connected components (SCCs) of a directed graph is a generic graph problem. For this purpose, we have proposed an implicit algorithm using BDDs. The method iteratively applies reachability analysis and computes SCCs sequentially. It has been used to compute the SCCs in the reachable state space of a set of finite state machines. Compared with an existing symbolic method which requires the transitive closure of the adjacency-matrix of the graph, our method is shown to be much faster, requires much less memory, and consequently solves much larger problems. Nevertheless, when there are excessive SCCs, our approach may not complete in a reasonable amount of time. Consequently, it might be interesting to explore a hybrid approach between our RA-based method and the TC-based method, which may be able to find multiple SCCs of relatively small size simultaneously but without computing the transitive closure of the entire graph.

We are currently studying two potential application of our method. One is for the symbolic home-state analysis of Petri nets (see, e.g., [11]). The other is the *bad-cycle-detection* problem in model checking [12, 13]. We expect that when the reachable state space has a small number of SCCs, our method may outperform the Emerson-Lei algorithm [12] and its refinements (e.g., [13]). The experiments for both applications are left as future work due to lack of time.

Acknowledgment

The authors would like to thank Fabio Somenzi of the University of Colorado at Boulder for insightful and motivating discussions regarding SCC identification.

References

[1] A. V. Aho, J. E. Hopcroft, and J. D. Ullman. *The Design and Analysis of Computer Algorithms.* Addison-Wesley Publishing Company, 1974.

[2] R. E. Tarjan. Depth first search and linear graph algorithms. *ACM Journal on Computing*, 1(2), 1972.

[3] R. E. Bryant. Graph-based algorithm for boolean function manipulation. *IEEE Transactions on Computers*, 35, August 1986.

[4] G. D. Hachtel, E. Macii, A. Pardo, and F. Somenzi. Markovian analysis of large finite state machines. *IEEE Transactions on Computer-Aided Design*, 15(12):1479–1493, December 1996.

[5] G. Hasteer, A. Mathur, and P. Banerjee. An implicit algorithm for finding steady states and its applications to FSM verifcation. In *Proc. ACM/IEEE Design Automation Conference*, pages 611–614, 1998.

[6] A. Xie and P. A. Beerel. Efficient state classification of finite state Markov chains. *IEEE Transactions on Computer-Aided Design*, 17(12):1334–1338, December 1998.

[7] V. Singhal. *Design Replacements for Sequential Circuits.* PhD thesis, University of California at Berkeley, 1996.

[8] S. Qadeer, R. K. Brayton, V. Singhal, and C. Pixley. Latch redundancy removal without global reset. In *Proc. International Conf. Computer Design (ICCD)*, Oct. 1996.

[9] J. R. Burch, E. M. Clarke, D. E. Long, K. L. McMillan, and D. L. Dill. Symbolic model checking for sequential circuit verification. *IEEE Transactions on Computer-Aided Design*, 13(4):401–424, April 1994.

[10] Brayton et al. VIS: A system for verfication and synthesis. In *Proc. International Conference on Computer Aided Verification*, pages 428–432, 1996.

[11] E. Pastor, J. Cortadella, and M. A. Peña. Structural methods to improve the symbolic analysis of Petri nets. In *Lecture Notes in Computer Science, LNCS 1639*, pages 26–45. Springer-Verlage, 1999.

[12] E. A. Emerson and C. L. Lei. Efficient model checking in fragments of the propositional modal mu-calculus. In *Proceedings of LICS 1986*, pages 267–278, 1986.

[13] R. H. Hardin, R. P. Kurshan, S. K. Shukla, and M. Y. Vardi. A new heuristic for bad cycle detection using BDDs. In *Proc. International Workshop on Computer Aided Verification*, pages 268–278, 1997.