Abstract
Software Testing is one among the major phases of Software Development Life Cycle (SDLC). Testing is a process carried out to ensure that the developed system is working as per the given requirements and is error-free [1] Test case generation is a major step in software testing. Software can be tested at any phase of SDLC. Earlier the system is tested, it becomes easier to reveal the faults and there by greatly reducing the cost and time required to develop the system. This paper aims at generating test cases from artefacts developed from design phase of the Software Development Life Cycle. UML Activity Diagram is selected as the design artefact to generate the test cases. The proposed approach generates optimized set of paths from the Activity Diagram. The main idea behind this is the efficiency of the testing process. Generating all possible test cases and then optimizing them is impractical. It will unnecessarily create either invalid test cases or redundant test cases and eventually makes the entire process of test case generation inefficient. This approach mainly aims at improving the efficiency of the system by automating the whole process and by generating only the optimized set of test cases.

Keywords: Model, Unified Modelling Language (UML), Activity Diagram, XML Metadata Interchange (XMI).

INTRODUCTION
Testing plays a major role in the development of a system. The testing process verifies whether the designed software works as per the user requirements. The testing process ingests nearly 80 percent of the cost of the software development. Testing can be done during any phase of the software development life cycle. The usual practice is to test the system after the implementation phase [1] In case of any defects at this phase, the entire system needs to be corrected to fix the defect. Therefore this process will increase the cost involved for the system development. This paper mainly aims at model based testing. In model based testing [14], design models are used to generate the test cases. With model based testing, it helps us to identify the faults at the early stages of the software development life cycle there by reducing the cost and time involved in the system development. Model is a graphical representation of the system functionality. In this paper, UML Activity Diagram is used as the design model to generate the test cases. UML is a Unified Modelling Language used to model the systems. UML has different notations and symbols used for representing different system functions and components.

The main purpose of this work is to analyze and automatically generate the effective set of test cases from a given activity diagram. A XMI parser is used to process the XML file being obtained from the UML activity diagram and maps it to its corresponding adjacency list. Then the test scenarios are generated by applying test data generation technique or by list traversal using AI techniques [21, 22].

The existing approaches provide a semi-automated way of test case generation from an UML activity diagram. That is, the given diagram is translated to some intermediate representation such as activity dependency graph, data flow graph, control flow graph or the activity dependency table. This intermediate representation is then given as input to the further phases. Even though it seems to be fully automated, the procedure for converting the diagram to its internal representation has not been addressed clearly in any of the existing approaches. The output of the system is optimized set of test suites and test cases. Without optimization the system would have generated all possible test suites out of which there could be even redundant or irrelevant test suites. Some test suites will have higher priority than others. And in case of Fork-Join activities the order of interleaving has to be maintained. Hence by taking care of these concepts, a Fork-Join coverage criterion has been applied to maintain the order of concurrent activities. The redundant paths are optimized using a Weighted Prioritized and Fork-Join coverage technique.

BACKGROUND
Activity Diagram is one of the behavioural diagrams in UML which describes the flow of activities in the system. Each activity can be further subdivided into smaller activities or it can be an atomic activity. The syntax of Activity Diagram is
very similar to that of flowcharts. Activity diagram mainly describes the way in which each functionality is executed in the system. It never explains “what the system does” rather explains “how the system does”. Activity Diagram uses variety of symbols to represent different scenarios. Every diagram will have one initial state and can have more than one final states. The start state in an Activity Diagram is denoted using a filled circle. Start state of an activity diagram will have no incoming edges. Activities are denoted using lozenge shape. If-Else scenario can be easily modelled by considering a decision construct which contains branches to represent true/false conditions. Diamond shaped symbol is used for decision making. UML also provides a mechanism to model concurrent activities of the system. When an activity needs to be run concurrently, we use Fork construct to model it. Fork is represented using horizontal thick line. Fork node contains one incoming edge which is split into multiple outgoing edges. The activities which are forked will execute concurrently. All the activities that are forked are joined later using join node. Join node is always used along with a Fork node. For an activity to occur in parallel it has to first fork and later once all the parallel activities are executed, they need to be joined together before starting with the next activity. End state of Activity Diagram is denoted using filled circle with an outer circle. End state will have no outgoing edges [2].

Fig. 1 shows a sample example of UAD consisting of seven activities. The various activities involved are A1, A2, A3, A4, A5, A6 and A7. Every Activity diagram contains only one start state and any number of end states. Start state in the Activity Diagram is denoted using filled circle. End state is shown using a filled circle with an outer circle. In the above example, after completion of activity A1, there is a fork node represented using thick horizontal line. The activities following fork node namely A2, A3, A4, A5 and A6 execute concurrently. The order of execution needs to be maintained among forked activities i.e. an activity can execute only when the execution of previous activity is complete. The If-Else construct is modelled using diamond symbol. Every decision node in an Activity diagram will have one incoming edge and can have two or more number of leaving edges. The decision symbol in Fig. 1 has two leaving transitions and each of them has a label referred as the guard expression [11, 15]. In Fig. 1 after executing the activity ‘A2’, if the condition of the branch node is evaluated to true, then ‘A5’ will be executed next, else the control moves on to activity ‘A4’. Merge nodes are also depicted in the same way as decision nodes but it acts as a source node to combine multiple flows. As per the example, activity A7 can begin only when all activities which were forked previously has completed their execution. A transition which is depicted as directed edge, describes the order of execution or control-flow between two activities. Consider two activities A and B, if there is a transition from A to B it means, Activity A has to be completed before Activity B starts.

![Figure 1. Sample Activity Diagram](image)

### LITERATURE SURVEY

Literature survey has been presented in table 1 to highlight the work that has been covered in this field earlier and also give the drawbacks of each work.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Description of method</th>
<th>Drawbacks identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Debasis Kundu and Debasis Samanta [6]</td>
<td>Activity diagram is first converted in to an Activity Graph using which the test cases are generated.</td>
<td>The paper has not clearly addressed the method to handle the fork-join constructs present in activity diagram.</td>
</tr>
<tr>
<td>A.Nayak and Debasis Samanta [4]</td>
<td>Based on the classification of control constructs followed by a transformation approach which takes into account any combination of nested structures and transforms an activity diagram into a model called Intermediate Testable Model (ITM). ITM was used to generate test cases.</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1: Survey of previous work
<table>
<thead>
<tr>
<th>Authors</th>
<th>Description of method</th>
<th>Drawbacks identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mitrabinda Ray et. al. [3]</td>
<td>This approach builds a flow dependence graph from UML activity diagram and applies conditioned slicing on a predicate node of the graph, in order to generate test cases.</td>
<td>Satisfies only path coverage criteria. Does not deal with activity and transition coverage.</td>
</tr>
<tr>
<td>Namita Khurana and R.S. Chollar</td>
<td>Generates test cases using sequence and state chart diagrams. The generated test cases are later optimized using genetic algorithm.</td>
<td>Some of the problem related to combination of Sequence and State Chart remain unresolved. Also the proposed approach is not yet automated.</td>
</tr>
<tr>
<td>Ajay Kumar Jena et al. [7]</td>
<td>Proposes an approach to generate test cases from UML 2.0 activity diagrams by covering maximum activity nodes in the diagram.</td>
<td>The approach is not automated. So an automated tool can be developed for the proposed approach.</td>
</tr>
<tr>
<td>Meliiana et al. [8]</td>
<td>The proposed approach modifies existing DFS algorithm to generate the test cases from design artifacts.</td>
<td>Contains redundant test cases.</td>
</tr>
<tr>
<td>Liping Lia et al. [9]</td>
<td>This paper combines software testing with Extenics and suggests an automatic approach for generating decreased set of test cases from UML Activity diagrams partly using Extension Theory.</td>
<td>Does not provide an efficient technique to handle fork and join constructs of an Activity Diagram.</td>
</tr>
<tr>
<td>Linzhang et al. [10]</td>
<td>In this approach, UML activity diagram is used to derive the test scenarios. Later, the information required for the generation of test cases is extracted from each test scenario and the possible values of all the input/output parameters could be generated by applying category-partition method.</td>
<td>Fails to handle the fork-join constructs present in activity diagram.</td>
</tr>
<tr>
<td>Chandler et al. [11]</td>
<td>The proposed architecture creates a special table called Activity Dependency Table (ADT) for each XML file using which it automatically generates a directed graph called</td>
<td>Does not handle the coverage criterions and fork join constructs of Activity Diagram.</td>
</tr>
<tr>
<td>Authors</td>
<td>Description of method</td>
<td>Drawbacks identified</td>
</tr>
<tr>
<td>-------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Activity Dependency Graph (ADG) that is used in conjunction with the ADT to extract all the possible final test cases.</td>
<td></td>
<td>Simple fork-join is considered.</td>
</tr>
<tr>
<td>Proposes a technique to prioritize test case scenarios by identifying the critical path clusters using genetic algorithm. The test case scenarios are derived from the UML activity diagram and state chart diagram.</td>
<td></td>
<td>The approach is semi-automated.</td>
</tr>
<tr>
<td>Proposes a prioritization technique based on UML activity diagrams. The constructs of an activity diagram are used to prioritize scenarios.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed an approach to generate test cases from activity diagrams by exposing all possible external inputs and outputs. It clearly explains the concept of test data generation at run time.</td>
<td></td>
<td>The techniques for handling loops have not been addressed in this paper.</td>
</tr>
<tr>
<td>In this approach, the given activity diagram is first converted into a grammar called as Activity Convert (AC) grammar. The generated test cases are then validated against the coverage analysis.</td>
<td></td>
<td>Redundant test cases are generated.</td>
</tr>
<tr>
<td>The paper mainly aims at producing transition paths considering the dependent transition pairs.</td>
<td></td>
<td>It does not consider other UML diagrams.</td>
</tr>
<tr>
<td>The paper aims at automatic generation of test cases from collaboration diagrams and confirms the result using mutation testing.</td>
<td></td>
<td>Not very feasible.</td>
</tr>
<tr>
<td>This paper presents an approach to prioritize the test scenarios using Genetic Algorithm along with Particle Swarm Optimization.</td>
<td></td>
<td>Not an automated approach.</td>
</tr>
<tr>
<td>A technique is proposed to generate test cases using mutated activity diagrams with the intention of finding defects.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
METHODOLOGY

Fig. 2, shows an overview of the proposed methodology. In the first step, activity diagram is designed using standard UML tool called StarUML. StarUML provides a workspace for the user to model different views of the system using Unified Modelling Language. It is not possible to perform any operations using this diagram, until and unless we have its internal representation. Hence the Activity diagram needs to be pre-processed. To achieve this, in the next step the diagram is exported to its corresponding XMI representation using StarUML. UML Activity diagram in its XML format consists of various XML tags and its attributes which are essential to produce the desired result. XMI file contains details about each of the activity and transitions in the activity diagram. Once the relevant details are obtained, the next step is to store this information into intermediate data structure (i.e. Adjacency list) for further processing. Adjacency list is traversed using AI techniques to obtain the optimized set of test scenarios. In the last step, test cases are generated from each of the scenario.

Several UML tools are available to model UML diagrams. Some of the widely used tools are MagicDraw, RationalRose, StarUML [12] etc. MagicDraw is a good suite for UML modelling. The tool manages a true model representation which it stores as one large XML flat file. The main reason for not using MagicDraw in the proposed approach is, the XML file generated by the tool for the diagram is very large and takes longer time for processing. RationalRose [13] is an object-oriented Unified Modelling Language (UML) software design tool intended for visual modelling and component construction of enterprise-level software applications. Rational Rose allows designers to take advantage of iterative development because the new application can be created in stages with the output of one iteration becoming the input to the next. The RationalRose tool doesn’t support option for exporting the diagram to its XMI/XML representation. Hence further processing of the diagram cannot be done using RationalRose. Both the tools mentioned above are proprietary software. StarUML is an open source tool which overcomes the above mentioned limitations and hence is preferred for the proposed approach.

Proposed methodology consists of following steps:

Step 1: Input to the system is an Activity diagram, which is modelled using StarUML tool. The Activity diagram considered for this project supports any number of connected activities, branching constructs and set of concurrent activities. By default, the diagram modelled using StarUML is saved in its .mdl format.

Step 2: Activity diagram in its .mdl format is converted to its XMI representation. XMI is a standard used to store UML models.

Step 3: The XMI file obtained in Step 2 is parsed to extract the required data. At this stage an algorithm is designed to parse the XMI file. Algorithm 1 explains the method used to extract the data from XMI file.

Algorithm 1 XMI Parser

procedure parseXMIStates( XMI file)
  Load the XML document
  if swim lanes are present then
    Check for the swim lanes in the diagram.
    Create a new list to hold all the swimlanes
    Add each activity to the swimlane list.
  else
    Add elements of compositeState.subvertex tag to sta for each element in sta.elements
    Retrieve elements name, id and kind.
    if incoming != NULL then
      Split the incoming edges based on ‘ ’
    else if outgoing != NULL then
      Split the outgoing edges based on ‘ ’
    end if
  end if
end procedure

Step 4: The XMI generated in Step 3 contains information about States and Transitions of Activity Diagram. The above algorithm makes use of two separate functions namely, parseXMLStates and parseXMLTransitions to parse states and its respective transitions. If the Activity diagram is modelled using Swimlanes, then the swimlane details will be enclosed within ActivityGraph.Partition element. The Activity state information is present within the CompositeState.subvertex element. Hence the file is traversed to get the descendants of CompositeState.subvertex element. In the first step, for each of the elements in CompositeState.subvertex, a state object is created, where all the attributes of state are initialized. For each of the element extract its id, name and kind. In the next step, the incoming transition is checked. If there is more than
one incoming transition in the Activity diagram then it is separated by a space in the XMI file which is later split based on space and stored in an array. The same procedure is then repeated for outgoing transitions. More than one outgoing transitions are split based on space and is stored in another separate array. In the third step, all the states of Activity diagram are added to the stateList. Once all the states of the Activity diagram have been parsed, next step is to parse the transitions of the diagram. Transition details are contained in UML::StateMachine.transition.

Step 5: Data retrieved in step 4 is populated into a data structure for further processing. For convenience, Adjacency matrix is of size N×N is used; where ‘N’ refers to number of activities in an Activity diagram. The entry A[i,j] in Adjacency matrix is set to transition name when there is a transition from Activity i to Activity j. The main reason for choosing adjacency list as a data structure is- each node in the list can be easily traversed to generate the set of test suites. Or every element in the above mentioned tag, a transition object is created. If the transition has any guard expression then the guard expressions attributes needs to be extracted. Once all the transitions of Activity diagram has been parsed, the transitions are added to transList.

Consider the Activity diagram in Fig. 3. The adjacency list representation for the same is shown in Fig. 4

Step 5: Once the data has been populated in Adjacency list, the list is traversed using AI technique called as Best First Search method (BFS). The BFS method is not used as it is; instead it is modified as per the need of the project. Best-first search in its most general form is a simple heuristic search algorithm. “Heuristic” here refers to a general problem-solving rule or set of rules that do not guarantee the best solution or even any solution, but serves as a useful guide for problem-solving. Best-first search is a graph-based search algorithm meaning that the search space can be represented as a series of nodes connected by paths.

The main objective here is to traverse the list and generate all possible paths from a Start activity node to a Final Activity node. The modified heuristic approach makes use of two lists namely Open List and Closed List. Open list is a list which contains the path without the Final node in it. Closed list is a list which contains the path with Final state. Every time a node is visited, it is first en-queued to the open path list. The last node in the list is then checked to see if it’s a final node. If it is a final node, then the whole path will be dequeued and will be added to Closed Path list. In other words, every closed path has been an open path in its lifetime. The same is explained using a pseudocode in Algorithm 2.

Algorithm 2: List_Traverse

Enqueue the start node in the openpaths Queue.
Repeat
Dequeue a path from the openpaths queue

Figure 3. Example Activity Diagram

Figure 4. Adjacency list representation
until openpaths queue is not empty
for each connected node n
    newpath ← oldpath + n
    if n is a final state then
        enqueue the newpath to closedpaths queue
    else
        enqueue the newpath to openpaths queue
    end if
end for

Path traversal in case of concurrent activities enclosed within fork-join pair. Path traversal in case of fork-join pair is a complex task. The activities within a fork-join pair can execute in an interleaved manner but, only condition is - an activity cannot start its execution until and unless the activity before it has completed its execution. As the number of activities within a fork-join pair increases; it results in path explosion and hence, is infeasible to consider all paths due to limited capability of resource and time. Most of the existing techniques generate all possible paths and then select the correct paths from the set of generated paths thereby eliminating the redundant and irrelevant paths. This will simply waste lot of time by generating irrelevant paths. Hence an attempt is made to reduce to the time involved in generating the possible paths. This concept is illustrated by considering few examples of activity diagram with fork-join constructs.

Step 6: Test case optimization - Test adequacy criteria are nothing but an essential and important predicate, which shows the adequacy. In other words it is a stopping condition to stop the testing process, when all the defined criteria have been met. Path coverage, transition coverage and activity coverage are few test adequacy criteria, used for the work.

Algorithm 3 Coverage Calculation
for each path in Final set of paths
    Create a new path object p
    Count total no. of states in the every path
    Count transitions in each path
    Assign weight for each of the path
    Weight = Weight + (inCounter*outCounter)
    \[ A_{Coverage} = \frac{\text{no of states Covered}}{\text{total states}} \times 100 \]
    \[ T_{coverage} = \frac{\text{no of transitions Covered}}{\text{total transitions}} \times 100 \]
end For

Algorithm 4 Fork-Join traversal
if node.kind == fork then
    Extract the corresponding join node
    FJTraversal (Start, End, Path)
end if

Algorithm 4 is applied to different activity diagram examples given in Fig. 5

Figure 5 (a) Two activities (b) Three activities (c) Four activities

The algorithm 4 is applied to various activity diagrams given under Fig. 5 and following paths are generated.

Start → A1 → A2 → A3 → A4 → A5 → A6 → End
Start → A1 → A2 → A4 → A3 → A5 → A6 → End
RESULT ANALYSIS

The proposed methodology is explained with a sample case study of ATM management system. The Automated Teller Machine will serve one customer at a time. A customer needs to first insert an ATM card and enter the four digit PIN code. Once the PIN is entered, the system will verify if the entered PIN is valid or invalid. If it is valid one, then it will proceed to the next step. If the entered PIN is an invalid one, then it will allow the user to enter the PIN again. For the sake of simplicity, the number of trails on an invalid input is set to one. (i.e. if the entered input is invalid, user will be given only once chance to re-enter the data). The ATM Management System supports. Once the user entered PIN is valid, in the next step system will ask the user to enter the choice of option. User can select one among the following choices:

- Withdrawal
- Balance Inquiry
- Mini Statement of the previous transactions

When the user clicks on Withdraw Option, the system will prompt the user to enter the amount to be withdrawn. Once the user enters, the amount entered will be verified to check if it is less than 500. If yes, the system will give a message saying insufficient balance and then asks the user to re-enter amount. Later, the amount value will be deducted from users account and cash will be dispensed to the user along with the receipt. When the user clicks on Balance Enquiry, the system will ask the user to confirm if he needs a printed receipt or no. If the user selects yes, then the system will output the balance details on a receipt and exits. If user selects no, then the system will output balance details on the screen and then exits. Mini Statement feature helps the user to know his past ten transaction details prior to current transaction. The Use case diagram for above mentioned case study is shown in Fig. 6 and Activity diagram for Withdraw Money use case is shown in Fig. 7.

Use Case diagram is one of the behavioral diagrams in UML. It is used to represent various users of the system and its various functionalities. In this ATM management system, there are two users namely Customer and Admin. Customer can withdraw money, enquire the balance or get the previous transaction details. Admin is the one who updates the ATM machine by periodically filling cash into it. The proposed methodology is tested on two functionalities namely Withdraw Money and Enquire Balance of the system. As a first step, Activity diagram is modelled for these two functionalities.
Fig. 8 shows various activities and its details obtained after parsing the XMI file. The first column in the table indicates the index of each activity. It is a unique number given to each activity, so that it becomes easy to access the activity whenever required. Second column gives the type of each node of an Activity diagram. As we have already seen earlier, there can be 4 kind of Activity node namely action state, pseudo state and final state. Third column gives the xmi id of each state. Fourth column of grid view gives name of each state of Activity Diagram. Next comes the kind of the node which can be either initial, branch, fork, join or final. This attribute is very much essential to traverse the path in case of fork-join pair or looping construct. Inc and Outg field in the grid view specify the incoming and outgoing transition ids of each state. Each transition is labelled as UMLTransition followed by a unique number. At last, there is a field called inCounter and outCounter, which keeps track of number of incoming and outgoing transitions. In total there are 14 states in the given AD starting from 0 to 13.

Fig. 9. Snapshot of the output showing various Test Paths for Withdraw Money Activity Diagram
Fig. 9 shows all possible test paths generated from the input activity diagram. The first three paths consider that the valid PIN is entered and permutes all three activities present with in fork-join pair. In the next three paths, it considers the entry of invalid PIN code which will cause the loop to be traversed again. Thus in total six paths are generated. For each of the paths its weight, activity coverage and transition coverage values are displayed. Weight of one node is given by: Number of incoming edges for a node X number of outgoing edges for a node. For a path, weight is sum of weights of nodes in that path. For the sake of simplicity, the activity coverage in case of loop is considered to be 100%.

For each of the test paths generated, a separate test case is generated. Since there are six different paths, we get six test cases shown in the snapshots shown below. Each test case consists of the state name in that path and test input and Expected Output for each of the states. The test cases generated here are not so accurate since starUML doesn’t support object diagram XMI generation. Hence the object diagram couldn’t be integrated with the activity diagram output.

The test case generated for one of the path of Withdraw Money AD is shown in Fig. 10.

<table>
<thead>
<tr>
<th>State Name</th>
<th>Test Input</th>
<th>Expected Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swipe Card</td>
<td>Card</td>
<td>Card details</td>
</tr>
<tr>
<td>Enter PIN</td>
<td>PIN code</td>
<td>Entered PIN code</td>
</tr>
<tr>
<td>Verify PIN</td>
<td>EntryAction1</td>
<td>Valid/Invalid PIN</td>
</tr>
<tr>
<td>Decision1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enter Withdrawal Option</td>
<td>Choice</td>
<td>Enter amount screen</td>
</tr>
<tr>
<td>Enter Amount</td>
<td>Amount value</td>
<td>a.Amount</td>
</tr>
<tr>
<td>Synchronization1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dispense Cash</td>
<td>a:amount</td>
<td>Cash</td>
</tr>
<tr>
<td>Deduct amount from Account</td>
<td>Amount value</td>
<td>Amount value is deducted from Account table</td>
</tr>
<tr>
<td>Prepare to print receipt</td>
<td>Transaction details</td>
<td>r:Receipt</td>
</tr>
<tr>
<td>Synchronization2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Display message</td>
<td>transaction complete</td>
<td>Display successful message</td>
</tr>
<tr>
<td>FinalState1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 10. Snapshot of the Sample Test case for Withdraw Money AD

CONCLUSION
Software Testing is a time consuming and costly process in software development life cycle. Automation of this phase may lead to overcome the above problems and also reduces the human effort in other ways it also helps in detecting the human intended errors and logical errors as well. Automation of testing will not be that much productive in terms of time consuming and cost, if we have to wait till the end of the SDLC stage i.e. if we follow the white box testing methodology of testing. If any errors will be detected in this stage, we have to go for that part of the code and design document as well. We have to follow up strict verification of both code and design document from beginning to short out the error. So only one solution to this problem is to, start the testing process from early stage of SDLC i.e. from requirement specification stage through design phase up to the last phase.

REFERENCES


[26] Xinying Wang, Xiajun Jiang and Huibin Shi, “Prioritization of Test Scenarios using Hybrid Genetic Algorithm Based on UML Activity Diagram”, 978-1-4799-8353-7/15/$31.00 ©2015 IEEE.