Secure UML Information Flow using FlowUML

Khaled Alghathbar¹, Duminda Wijesekera¹, and Csilla Farkas²†

¹ Dept. of Information and Software Engineering and CSIS, George Mason University, MS 4A4, Fairfax, VA 22030
² Information Security Laboratory, Dept. of Computer Science and Engineering, University of South Carolina, Columbia, SC 29208

Abstract. FlowUML is a logic-based system to validate information flow policies at the requirements specification phase of UML based designs. It uses Horn clauses to specify information flow policies that can be checked against flow information extracted from UML sequence diagrams. FlowUML policies can be written at a coarse grain level of caller-callee relationships or at a finer level involving passed attributes.

1 Introduction

As security becomes an important aspect of large software systems, it needs to be addressed throughout the software development life cycle. Considered a non-functional requirement, security requirements have not been considered during the early phases of system development; thus resulting in vulnerable software [7]. In [13,14] Nuseibeh and Easterbrook present the need to formally analyze and validate security requirements during earlier phases to detect and remove design vulnerabilities. Recent work has addressed information flow control security [3,1,12,16]. Information flow requirements need to be developed and evaluated during the requirements and design stages of the software life cycle, as validating information flow requirements at an early stage prevents costly fixes mandated during latter stages of the development life cycle. In this paper we propose a logic-based system (FlowUML) to validate information flow policies at the requirements specification phase of UML-based designs [17]. FlowUML uses locally stratified Horn clauses to enforce user specifiable information flow policies via three processes: 1) Extract information flows predicates from UML sequence diagram, 2) Derive all inherited and indirect flows, and 3) Check for compliances with specified policies - specified at two levels of granularity. At the coarse level, a flow is represented as a directed arc between components, going from its source to its sink. At the fine level, the flow is defined based on the semantics of methods, attributes passed between the components, and the roles played by the components.

† Farkas’ work was partially supported by the National Science Foundation under grant number IIS-0237782.
FlowUML does not assume the existence of a particular meta-policy and can support a variety of policies. FlowUML advances over existing secure information flow models by applying a formal framework to the early phases of the software development life cycle. Towards that end, Section 2 shows information flow specifications embedded in UML sequence diagrams. Section 3 provides a running example. Section 4 describes our notations and assumptions. Section 5 explains the FlowUML process. Section 6 has the syntax and semantics of FlowUML. Section 7 shows the expressibility of FlowUML by means of the running example. Section 8 describes related work and Section 9 concludes the paper.

2 Flow Specification in UML

The UML is the de-facto design language for large software projects. The UML uses multiple diagrams (views) to represent requirements and designs. Use Cases support the specification of usage scenarios between the system and its intended users [17]. Interaction diagrams specify how objects in the use case interact with each other. Sequence diagrams are specific interaction diagrams that show such interactions as a sequence of messages exchanged between objects over time. The sequence diagrams shown in Fig. 1 consists of four objects and two actors, where Actor A initiates the first flow as a request to read information from Obj3 which then returns the requested information. Secondly, Actor A writes information to Obj4 and the control object. The flow to the control object triggers additional flows. FlowUML extracts flows from the sequence diagrams.

![Fig. 1. A sample sequence diagram case1](image1)

Fig. 2. The sequence diagram for use case1

3 Running Example

Our example consists of two use cases. Use case 1 has a simple scenario shown in Fig. 2, and use case 2 has a more complex scenario shown in Fig. 3. The actor in use case 1 reads information from object Obj1 returned as an attribute att1 and then writes it to Obj2. The actor in use case 2 transfers information between two objects and then calls another object leading to additional information flow.

During the analysis stage of the software development, selected interactions are specified between objects. These interactions are further refined by specifying the
message’s attributes or parameters passed between objects. Therefore, the expressiveness and detail in a sequence diagram depends on the analysis stage. For example, the sequence diagram of use case 2 can be categorized into two views: coarse grain view shown in Fig. 1 and the fine grain view shown in Fig. 3.

Flow and access control requirements are expressed over the actors and objects of the sequence diagram. For example, we may specify that information is permitted to flow from Obj 3 to Obj 4 but not vice versa. Similarly, we may permit or deny Actor A to read Obj 3.

![Flow and Access Control Requirements](image)

**Fig. 3. Detailed sequence diagram for use case 2**

## 4 Notations and Assumptions

Objects in sequence diagrams belong to three categories: actors, entities, or control objects. Actors initiate flows and send/receive information to/from the system as externals. Every use case has at least one actor. Entity objects are persistent and store information. Control objects provide coordination and application logic. Generally, actors and entity objects can be sources or sinks of information flows and control objects do not store information. However, they may create new information without storing them. This is a basic axiom of FlowUML.

**Axiom 1:** Sources and sinks of information flows are actors and entities.

Information flows in sequence diagrams arise due to attributes passed in method calls such as read and write. For example, the method `read(att1)` exchanges information between Obj 3 and Actor A. FlowUML uses any message as a flow of information regardless of its name.

**Axiom 2:** Any information-carrying message is considered an information flow.

Sequence diagram builds complex messages from simple ones using three kinds of constructs: **creation messages** used to create new objects, **iteration messages** used to send data multiple times, and **conditional messages** used to control messages based on satisfying certain conditions. We consider a creation message as an information flow if the caller passes creation parameters to the callee. We consider an iterated message as a single information flow. We consider a conditional message to be an information flow regardless of the truth-value of the condition. We consider a simple message as a flow if it passes information.
4.1 Attribute Dependencies across Objects

We make the following observations about the attributes:
1. Information flowing into an object may flow out unaltered, called exact attribute flow in FlowUML.
2. Some attributes flowing out of an object may be different in name but always have the same value as an attribute that flow into the same object. FlowUML requires this information to be in the similar attribute table.
3. Some attributes flowing out of an object may depend upon others that flow into the same object. FlowUML requires this information to be in the derived attribute table.

We formalize these observations in the following definitions and axioms.

**Definition 1:** An exact attribute is one that flows through an object but does not change its value. An attribute that flows out of an object depends upon a set of attributes flowing into the same object. We say that the output attribute is derived from the input attributes.

**Axiom 3:** Attribute names are unique over a system.

**Axiom 4:** Pairs of exact attributes are entered into the exact attribute table and dependent attributes have an entry in the derived attribute table.

At the requirements specification stage flow analysis depends on 1) the types of components (i.e., actor, entity or control object), 2) the sequence of objects in a flow, and 3) whether the flow is within a use case or between many use cases. For example, entities or actors may alter information while control objects do not.

5 The Verification Process

FlowUML verify flow policies using five steps using two metadata sources as shown in Fig. 4. Coarse grain policies require less detail than fine grain policies. Fine grain policies incorporate attribute values and their dependencies.
Additional Metadata for the fine-grain policy

Metadata
Flow Structure  Flow Constraints

Metadata
Additional Metadata for the fine-grain policy
Similar Attributes  Derived Attributes

Fig. 4. Steps in FlowUML

Coarse grain policy analysis

In the first step FlowUML extracts flow structure, called the defining flow structure, from Sequence diagrams and transforms it into basic predicates (see Section 6). The second step propagates basic predicates using the actor (or role) hierarchy, deriving implied information flows. The third step derives all transitive flows. The fourth step complements the third step by filtering results of the third step to those flows that satisfy properties of interests as specified in the policies. For example, in Fig. 1 the third step derives a flow from Obj3 to Actor A directly, flow from Obj3 to Obj4 through Actor A, flow from Obj3 to control through Actor A etc. In the fourth step, predefined policies may only apply to non-transitive flows between only entities and actors. Thus only flows from Obj3 to Actor A and the one from Actor A to Obj4, but not the flow from Obj3 to Obj4 may be of relevance. The fifth step detects flows that violate specified policies. Currently, FlowUML does not attempt to resolve the detected conflicts automatically.

Fine grain policy analysis

Fine grain policies require two kinds of additional information. The first, given in the similar attribute table, contains name of the attributes that always contain the same data value. The second, given in the attribute derivation table, lists attribute dependencies. This information is useful in controlling the flow of sensitive information. For example, in Fig. 3, if there is a record in the attribute derivation table stating that att6 is derived from att3 in the control object then FlowUML concludes that there is a transitive flow from Actor A to Actor B.

6 Syntax and semantics

FlowUML alphabet contains variables or constants over actors (A), objects (Obj), use cases (UC), attributes (Att), and times (T). Constants are members of the sets A, Obj,
UC, Att, and T. Variables are represented as $X_a$, $X_{obj}$, $X_{uc}$, $X_{att}$ and $X_t$, respectively.

FlowUML uses the following predicates:

**Basic predicates for supporting the specification of flows, called FlowSup.**

1. A unary predicate isEntity($X_{obj}$) meaning $X_{obj}$ is an entity.
2. A unary predicate isActor($X_{obj}$) meaning $X_{obj}$ is an actor.
3. A binary predicate specializedActor($X_a$, $X'_a$) meaning $X'_a$ is a specialized actor of $X_a$.
4. A binary predicate precedes($X_{uc}$, $X'_{uc}$) meaning use cases $X_{uc}$ and $X'_{uc}$ are executed in that order.
5. A binary predicate sameAtt($X_{att}$, $X'_{att}$) meaning attributes $X_{att}$ and $X'_{att}$ have the same value.
6. A binary predicate derAtt($X_{att}$, $X'_{att}$, $X_{obj}$) meaning that attribute $X'_{att}$ is derived from attribute $X_{att}$ in object $X_{obj}$.
7. A ternary predicate ignoreFlow($X_a$, $X_{obj}$, $X'_{obj}$) meaning to exclude the flow from object $X_{obj}$ to object $X'_{obj}$.
8. A 6-ary predicate ignoreFlows($X_{a1}$, $X_{obj1}$, $X'_{obj1}$, $X_{a2}$, $X_{obj2}$, $X'_{obj2}$) to exclude the flow from $X_{obj1}$ to $X'_{obj}$ and the flow from $X_{obj2}$ to $X'_{obj2}$ from being detected as a violation of flow control policies. Two similar predicates ignoreFlows($X_{a1}$, $X_{att}$, $X_{obj1}$, $X'_{obj1}$, $X_{a2}$, $X_{obj2}$, $X'_{obj2}$) and ignoreFlow($X_{a1}$, $X_{att}$, $X_{obj1}$, $X'_{obj1}$, $X_{a2}$, $X_{obj2}$, $X'_{obj2}$) are used in fine grain policies.

**Predicates for specifying flow constraints, called ConstSup.**

1. A binary predicate dominates($X_{obj}$, $X'_{obj}$) meaning that the security label of object $X'_{obj}$ dominates or equals the security label of object $X_{obj}$. It is used in multi-level security policies.
2. A binary predicate ACL($X_{obj}$, $X'_{obj}$, $X_{AT}$) meaning that object $X_{obj}$ is in the access control list of object $X'_{obj}$. It is used in discretionary access control policies. $X_{AT}$ is an operation such as read or write.
3. A binary predicate conflictingActors($X_{obj}$, $X'_{obj}$) meaning that actors $X_{obj}$ and $X'_{obj}$ are in conflict with each other for a specific reason. For example, both actors can not flow information to the same object or there must not be a flow of information between them.
4. A binary predicate conflictingEntities($X_{obj}$, $X'_{obj}$) meaning both entities $X_{obj}$ and $X'_{obj}$ are in conflict with each specific reason. For example, each entity belongs to different competitive company and information must not flow in between.

**Predicates for coarse-grain policies**

1. A 5-ary predicate Flow($X_a$, $X_{obj}$, $X'_{obj}$, $X_t$, $X_{uc}$) meaning there is a simple flow initiated by actor $X_a$ from object $X_{obj}$ to object $X'_{obj}$ at time $X_t$ in use case $X_{uc}$.
2. A 5-ary predicate mayFlow($X_a$, $X_{obj}$, $X'_{obj}$, $X_t$, $X_{uc}$) is the transitive closure of the previous predicate.
3. mayFlowInterUC($X_a$, $X_{obj}$, $X'_{obj}$, $X_{uc}$) is similar to mayFlow but the scope of mayFlowInterUC is between use cases instead of focusing in one use case.
4. A 4-ary predicate finalFlow($X_a$, $X_{obj}$, $X'_{obj}$, $X_t$) meaning a finalized flow.
5. finalFlowInterUC($X_a$, $X_{obj}$, $X'_{obj}$, $X_{uc}$) is similar to finalFlow but it covers flows between use cases.
6. A ternary predicate unsafeFlow($X_a$, $X_{obj}$, $X'_{obj}$) meaning there is an unsafe flow from $X_{obj}$ to $X'_{obj}$ initiated by actor $X_a$. 

7. A 6-ary predicate \( \text{unsafeFlows}(X_{a1}, X_{obj1}, X'_{obj1}, X_{obj2}, X'_{obj2}) \) meaning there are two unsafe flow. The first, initiated by actor \( X_{a1} \) flows from \( X_{obj1} \) to \( X'_{obj1} \). The second, initiated by actor \( X_{a2} \) flows from object \( X_{obj2} \) to object \( X'_{obj2} \). They are unsafe because together they violate a flow constraint.

8. A ternary predicate \( \text{safeFlow}(X_{a}, X_{obj}, X'_{obj}) \) meaning that the flow initiated by actor \( X_{a} \) from object \( X_{obj} \) to object \( X'_{obj} \) is safe.

**Predicates for fine-grain policies**

The predicates used to specify fine grain access control policies are similar to the ones for coarse grain policy, but include \( X_{att} \) as an attribute flowing between objects. The predicates are as follows: \( \text{flow}(X_{a}, X_{att}, X_{obj}, X'_{obj}, X_{t}, X_{uc}) \), \( \text{mayFlow}(X_{a}, X_{att}, X_{obj}, X'_{obj}, X_{t}, X_{uc}) \), \( \text{mayFlow}_{\text{interUC}}(X_{a}, X_{att}, X_{obj}, X'_{obj}, X_{t}, X_{uc}) \), \( \text{finalFlow}(X_{a}, X_{att}, X_{obj}, X'_{obj}, X_{t}) \), \( \text{finalFlow}_{\text{interUC}}(X_{a}, X_{att}, X_{obj}, X'_{obj}, X_{t}) \), \( \text{unsafeFlow}(X_{a}, X_{att}, X_{obj}, X'_{obj}) \), \( \text{unsafeFlows}(X_{a1}, X_{att1}, X_{obj1}, X'_{obj1}, X_{a2}, X_{att2}, X_{obj2}, X'_{obj2}) \), \( \text{safeFlow}(X_{a}, X_{att1}, X_{obj}, X'_{obj}) \).

FlowUML uses a set of predicates, as summarized in Tables 1 and 2. A FlowUML rule is of the form \( L \leftarrow L_1, \ldots, L_n \) where \( L, L_1, \ldots, L_n \) are literals satisfying the conditions stated in Tables 1 and 2. Rules constructed according to these specifications form a locally stratified logic program, and therefore has a unique stable model and that stable model is also a well-founded model [11].

**Table 1. FlowUML’s strata for coarse grain policies**

<table>
<thead>
<tr>
<th>Phase</th>
<th>Stratum</th>
<th>Predicate</th>
<th>Rules defining the predicate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coarse-grain</td>
<td>0</td>
<td>FlowSup predicates</td>
<td>base relations.</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Flow(X_a, X_obj, X'_obj, X_t, X_uc)</td>
<td>body may contain FlowSup predicates.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>mayFlow(X_a, X_obj, X'_obj, X_t, X_uc)</td>
<td>body may contain literal from strata 0 to 2.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>mayFlow_{\text{interUC}}(X_a, X_obj, X'_obj, X_t, X_uc)</td>
<td>body may contain literal from strata 0, 1 and 3.</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>finalFlow(X_a, X_obj, X'_obj, X_t)</td>
<td>body may contain literal from strata 0 to 3.</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>unsafeFlow(X_a, X_obj, X'_obj)</td>
<td>body may contain literal from strata 0 to 4.</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>safeFlow(X_a, X_obj, X'_obj)</td>
<td>body may contain literal from strata 0 to 5.</td>
</tr>
</tbody>
</table>

**Table 2. FlowUML’s strata for fine grain policies**

<table>
<thead>
<tr>
<th>Phase</th>
<th>Stratum</th>
<th>Predicate</th>
<th>Rules defining the predicate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fine-grain</td>
<td>0</td>
<td>FlowSup predicates</td>
<td>base relations.</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Flow(X_a, X_att, X_obj, X'_obj, X_t)</td>
<td>body may contain FlowSup predicates.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>mayFlow(X_a, X_att, X_obj, X'_obj, X_t)</td>
<td>body may contain literal from strata 0 to 2.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>mayFlow_{\text{interUC}}(X_a, X_att, X_obj, X'_obj, X_t)</td>
<td>body may contain literal from strata 0, 1 and 3.</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>finalFlow(X_a, X_att, X_obj, X'_obj, X_t)</td>
<td>body may contain literal from strata 0 to 3.</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>unsafeFlow_{\text{att}}(X_a, X_att, X_obj, X'_obj)</td>
<td>body may contain literal from strata 0 to 4.</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>safeFlow_{\text{att}}(X_a, X_att1, X_obj, X'_obj)</td>
<td>body may contain literal from strata 0 to 5.</td>
</tr>
</tbody>
</table>
7 Applying FlowUML

This section shows how some sample FlowUML policies and apply them to detect unsafe flows. Due to space limitation, we only present the basic flow predicates, propagation policies, and the security verification for discretionary access control.

**Basic flow predicates**: Examples flow information available in Fig. 2 are given in rules (1). The rules (2) to (6) are instances of FlowSup predicates of Figs 2 and 3.

```
Flow_att(A, att1, Obj1, A, Obj2, 2, uc1) ← Flow_att(A, att1, A, Obj2, 2, uc1) ← (1)

isEntity(obj1) ←, isActor(actorA) ←, precedes(uc1, uc2) ← (2, 3, 4)

sameAtt(att3, att3') ←, derAtt(att3, att6, control) ← (5, 6)
```

**Propagation policies**: The second step applies a policy that propagates flows along actor hierarchies. In our example, if there is a specialized actor say Actor C of Actor B then Actor C receives att5 and att6. Policies stating acceptable inheritances can be stated in example rules such as (7) to (9).

```
Flow(X', a, X, obj, X', obj, X, t, X, uc) ← Flow(X, a, X, obj, X', obj, X, t, X, uc), specializedRole(X, X') (7)

Flow(X, a, X', obj, X, obj, X, uc) ← Flow(X, a, X, obj, X', obj, X, uc), isActor(X, obj), specializedActor(X, obj, X' obj) (8)

Flow(X, a, X, obj, X', obj, X, uc) ← Flow(X, a, X, obj, X', obj, X, uc), isActor(X', obj), specializedActor(X', obj, X' obj) (9)
```

**Detecting flows unsafe with respect to policies**: This section shows FlowUML specification of known flow control policies, and how to detect information flows unsafe with respect to them.

**Mandatory Access Control (MAC)** permits a subjects to read an objects only if the subject’s security clearance dominates the security classification of the object (2).

Rule 18 specify if information flows from Obj1 to Obj2. If Obj2 does not dominate or equals the security label of Obj1 then the flow is considered unsafe.

```
unsafeFlow(X, a, X, att1, X, obj1, X, obj2, X, t) ← finalFlow(X, a, X, att1, X, obj1, X, t), ¬dominates(X, att1, X, obj2) (18)
```

**Discretionary access control (DAC)** allows subjects to access objects solely based on the subject’s identity and the authorization rule. Rule 19 specify unauthorized information flows from an actor to an object and rule 20 specify unauthorized flows from an object to an actor.

```
unsafeFlow(X, a, X, att1, X, obj1, X, obj2, X, t) ← finalFlow(X, a, X, att1, X, obj1, X, t), ¬ACL(X, att1, X, obj2, w) (19)

unsafeFlow(X, a, X, att1, X, obj2, X, obj1, X, t) ← finalFlow(X, a, X, att1, X, obj2, X, t), ¬ACL(X, att1, X, obj1, r) (20)
```
8 Related Work

Secure information flow models have been studied by several researchers. Most of these develop richer flow control models. For example, Samarati et al. develops a DAC based model that prevents information leakage by Trojans [16]. Bertino et al. [3] present a logic based workflow model. Although important, these papers do not address flow policies for software design life cycle models.

Myers [12] presents JFlow, an extension to Java that adds statically checkable flow constraints. However, JFlow can be used during the implementation phase while FlowUML is to be used during the requirements, design and analysis phases of the software development life cycle. Lodderstedt et al. [18] present a language to annotate UML-based models representing authorization constraints. Jurjens [19] proposes to extend UML, allowing to represent security information within UML diagrams.

To the best of our knowledge, none of the existing works address the design and verification processes supported by FlowUML. However, some published work shares some of FlowUML’s objectives. In the area of logic-based checking of policy violations during the requirement engineering, Alghathbar et al. [1] developed AuthUML, a logic program based framework for that analyzing access control requirements during the requirements engineering phase to ensure consistency, completeness and conflict-freedom. AuthUML is the access control analog of FlowUML.

FlexFlow of Chen et al. [6] is a logic-based flexible flow control framework to specify data-flow, workflow and transaction systems. Although FlowUML and FlexFlow analyze and prevent unauthorized flows, FlowUML is different from FlexFlow in many aspects such as FlexFlow is an abstract model while FlowUML is tightly integrated with UML diagrams.

9 Conclusions

FlowUML is a logic programming based framework to specify and validate information flow policies in UML based designs at the early phases of the software development life cycle. We demonstrated the expressiveness of FlowUML by validating existing example policies.

References


