Testing Context-Aware Middleware-Centric Programs: a Data Flow Approach and an RFID-Based Experimentation [*†]

Heng Lu
The University of Hong Kong
Pokfulam
Hong Kong
hlu@cs.hku.hk

W. K. Chan
City University of Hong Kong
Tat Chee Avenue
Hong Kong
wkchan@cs.cityu.edu.hk

T. H. Tse
The University of Hong Kong
Pokfulam
Hong Kong
thtse@cs.hku.hk

ABSTRACT

Pervasive context-aware software is an emerging kind of application. Many of these systems register parts of their context-aware logic in the middleware. On the other hand, most conventional testing techniques do not consider such kind of application logic. This paper proposes a novel family of testing criteria to measure the comprehensiveness of their test sets. It stems from context-aware data flow information. Firstly, it studies the evolution of contexts, which are environmental information relevant to an application program. It then proposes context-aware data flow associations and testing criteria. Corresponding algorithms are given. It uses a prototype testing tool to conduct experimentation on an RFID-based location sensing software running on top of context-aware middleware. The experimental results show that our approach is applicable, effective, and promising.

Categories and Subject Descriptors

D.2.5 [**Software Engineering**]: Testing and Debugging—*Testing tools*; D.2.8 [**Software Engineering**]: Metrics—*Product metrics*

General Terms

Experimentation, Verification

Keywords

context-aware applications, RFID, test adequacy

1. INTRODUCTION

Radio Frequency Identification or RFID is widely considered as an enabling technology ranging from Internet payment systems to supply chain management. For example, Wal-Mart in USA, Metro in Europe, and Hutchison Ports in Asia are implementing their RFID solutions. "These companies were attracted to RFID because it held out the potential of offering perfect supply chain visibility — the ability to know the precise location of any product anywhere in the supply chain at any time." [1] Because of the emerging importance of RFID technology, the testing of their software applications is essential. [2] Since precise location tracking of products (such as within an area of one square meter) is desirable and yet unattainable [24], it is difficult to determine whether the imprecision is caused by a fault in the software, hardware, or both. Moreover, although researchers [35] are tackling the test oracle problem for the software part, little progress has been reported in the literature on software test adequacy, "a criterion that defines what constitutes an adequate test" and one of the most important research problems in software testing [40].

In this paper, we shall focus on the study of the measurement problem in software test adequacy for context-aware applications. We propose a novel family of context-aware data flow test adequacy measurement criteria for context-aware middleware-centric applications. We apply and evaluate our proposal in an RFID-based location sensing program on a context-aware middleware [37, 38] with a published estimation algorithm [24]. The experimentation shows that our approach is very promising.

Pervasive computing [5, 28, 35, 39] has two core properties, namely context awareness and ad hoc communication. The context-aware property enables an application to acquire and reason about environmental attributes known as *contexts*. Based on the contextual information, the application can react impromptu to changes in the environment. The ad hoc communication property facilitates mobile interactions among components of the application with respect to the changing contexts. For example, a "polite" phone for car drivers by Motorola Labs will transfer an ongoing call to the speaker phone when entering a car park, route calls to a voice mailbox in complex driving situations, and call a pre-defined emergence number if an airbag has been deployed. [3]

Many existing proposals for pervasive computing are middleware-centric [3, 4, 5, 17, 28, 39]. One of the major reasons is that the middleware-centric multi-tier architecture favors the development and configuration of pervasive computing applications in terms of the *separation of concern* in a highly dynamic and compositional environment, in which the middleware transparently

[*] This work is supported in part by a grant of the Research Grants Council of Hong Kong (project no. HKU 7175/06E) and a grant of The University of Hong Kong.

[†] Part of the work was done when Chan was with The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong.

[1] Source: "What is RFID", *RFID Journal*, available at http://www.rfidjournal.com/article/articleview/1339/1/129/.

[2] See, for example, the scope of Sun's RFID Test Center in Dallas, Texas available at http://www.sun.com/software/solutions/rfid/testcenter/.

[3] Source: R. M. Gardner, "Technology solutions toward improved driver focus", Panel on Technology and Distracted Driving, International Conference on Distracted Driving, Toronto, Canada, October 2–5, 2005.

acquires, disseminates, and infers the contexts on behalf of the applications over ad hoc networks. For the purpose of brevity, we shall limit ourselves in this paper to the study of context-aware middleware-centric systems. We shall refer to a context-aware middleware-centric program simply as a *CM-centric program*.

A component of an application hence interacts with the middleware or other components of the application via a clear, loosely coupled and context-aware interface residing in the middleware. Although context-aware middleware can handle the ad hoc communication property, the other core property — the context-aware property — requires the collaboration of application components; otherwise, an application may not be context-aware. In the rest of the paper, an application function directly invoked by a middleware will be referred to as an *adaptive action*.

Because of such a common design in context-aware systems, the program logic of a CM-centric program normally spans over the application tier and the middleware tier. Firstly, a context-aware interface at the middleware tier may invoke an adaptive action whenever the middleware detects the interesting contexts registered in the interfaces by the CM-centric program. Secondly, the invoked adaptive action would serve as an entry of the CM-centric program at the application tier to react to the interesting contexts. Thirdly, other actions of the CM-centric program will utilize the results of the adaptive actions to provide tailored services to its users. Since black box testing techniques do not consider the structural organization of program units in a context-aware system with middleware support, they are intuitively less effective than their white box counterparts in detecting faults specific to context-aware systems designed by the above approach. We shall, therefore, restrict ourselves in this paper to white-box testing techniques.

The structural organization of program logic poses challenges to white-box testing of CM-centric programs, since it would be inadequate to consider the program structure of the components in the application tier alone. As to be explained in Section 3, a fault in the context-aware interface may reduce the efficacy of the context-awareness and is hard to be revealed by traditional approaches, which only consider the adaptive actions. Thus, it is necessary to enhance existing test adequacy criteria.

The main contributions of the paper are as follows: (*i*) This paper is among the first of its kind — we are not aware of any work in the literature that takes the context-aware middleware into account when formulating test adequacy criteria. (*ii*) New types of data flow associations are formalized to capture the context-aware dependencies specific to pervasive CM-centric programs. (*iii*) A novel family of data flow test adequacy criteria is proposed to measure the quality of test sets. The corresponding algorithms are designed accordingly.

The proposed family of adequacy criteria is evaluated on the *Cabot* platform, a pervasive context-aware middleware system [37, 38]. Our prototype testing tool automatically generates adequacy test sets according to our family of adequacy criteria. Evaluation of the fault-detection rates of these test sets indicates that our approach is effective and promising.

The rest of the paper is organized as follows: Sections 2 and 3 outline the technical preliminaries and testing challenges of CM-centric programs, respectively. Next, Section 4 will present our novel data flow associations followed by our testing criteria to measure the test sets comprehensiveness in Section 5. Section 6 evaluates our proposal by an RFID-based experimentation. This is followed by discussions, a review of related work, and the conclusion in Sections 7, 8, and 9.

2. CM-CENTRIC PROGRAMS

2.1 Fundamentals

In this section, we formally explain the fundamentals on which our testing proposal relies. We formalize our model based on our understanding of the common designs in contemporary CM-centric projects and technologies [3, 4, 5, 17, 28, 39].

A *context* of an entity characterizes its environmental attributes. An entity can be of diverse granularities such as a composite component, an object, or a pattern. We follow the popular tuple space representation in defining contexts [6]. We shall assume that C denotes the set of all context variables and V_c denotes the domain of values applicable to the context variable c.

DEFINITION 1 (CONTEXTS). *A* **context** *is an ordered couple* $\langle c, v\rangle$*, where c is a context variable and v is a value in* V_c.

The middleware continually detects all the context changes by assigning an instance value v to a context variable c [38]. Such detections and instantiations are transparent to applications atop the middleware. Since every context variable is unique in a CM-centric program, any occurrence of a context variable in the program will consistently give the same context evolution. Moreover, as we shall show later, the key-value model is sufficient to capture the interesting data flow information among contexts. To maintain generality, therefore, we shall refrain from using a more specific model.

Next, we formalize the notion of context-aware interaction between a middleware and an application. It is a common design that the middleware tier contains a context reasoning component, whose applications can define a set of rules, known as *situations*, to describe the interesting conditions over context variables [5, 6, 26, 37, 39]. When an interesting condition is satisfied for a particular combination of values of context variables, an adaptive action in the application will be invoked mechanically by the middleware. We assume that the middleware uses the standard event-condition-action (ECA) approach (as in [1, 22], for instance) to invoke the adaptive actions. In the rest of the paper, we refer to the above process as a *context-aware adaptation*.

Based on this understanding of context-aware adaptation, we formulate the concept of a situation as follows:

DEFINITION 2 (SITUATIONS). *A* **situation** *is a triple* $\langle C_s, p, Act\rangle$*, where* C_s *is a set of context variables that the situation subscribes, p is a triggering condition whose variables are in* C_s*, and Act is an adaptive action to be invoked by the middleware if and only if p is evaluated to be true, denoted by* $p(C_s) = true$*. A situation is said to be* **satisfied** *if* $p(C_s) = true$*; otherwise it is said to be* **outstanding**.

To enable context awareness, a context reasoning component at the middleware will receive the updates of the subscribed context variables from the context detection component [37, 38, 39]. When a situation is satisfied, the middleware will invoke an adaptive action. We assume that such a mechanism is spontaneous. Without loss of generality, each situation will be bound to at most one adaptive action, as in Definition 2.[4] Before we introduce the motivation example, we formally define CM-centric programs as follows:

DEFINITION 3 (CM-CENTRIC PROGRAMS). *A* **Context-Aware Middleware-Centric Program***, or* **CM-centric program**

[4] A situation involving multiple actions can be considered as a set of situations, each of which is associated with one adaptive action.

for short, is a triple $\langle C, U, S\rangle$, where C is a set of context variables, U is a set of adaptive actions, and S is a set of situations such that, for every situation $s = \langle C_s,\, p, Act\rangle \in S$, $Act \in U$ and $C_s \subseteq C$.

When a standard program is seen as a set of program units, a CM-centric program extends it by taking also its context-aware interfaces into account. Hence, we assume that all interactions of a CM-centric program with its environment are made through the context-aware interfaces.

2.2 A Motivation Example

In this section, we outline a fragment of a sample CM-centric program, which is a smart streetlight application implemented on the *Cabot* middleware platform. The application scenario [35] is as follows:

> Consider a system of smart streetlights that collaborate to illuminate a city zone. It includes two features. (*i*) Every visitor can personalize their favorite level of illumination irrespectively of their location within the zone. (*ii*) At the same time, the system maximizes energy savings by dimming unnecessary streetlights. When there is no visitor nearby, a streetlight will turn itself off. When a visitor walks toward a particular streetlight, the light detects the visitor and brightens itself. A streetlight nearby may dim itself if the closest light has provided sufficient illumination. Another streetlight may not dim, however, if there are other visitors requiring illumination. Because of the interference from other light sources and the presence of other visitors nearby, the resulting illumination for the visitor may differ from their favorite level. Finally, the system assumes that the effective distance for any streetlight to serve a visitor is at most 5 meters.

Suppose the application defines two situations, namely a visitor appearing nearby ($s_{visitor\_nearby}$) and the illuminance being lower than the favorite level ($s_{low\_illuminance}$). Their implementations are shown in Tables 3 and 4, respectively, in a tabular format for the ease of understanding. Take the situation *low_illuminance* in Table 4 as an example. The middleware defines $s_{low\_illuminance}$ as a triple $\langle C_{low}, p_{low}, incCurrent\rangle$. C_{low} contains all the context variables that will be referenced or updated in the triggering condition and/or the adaptive action. The set of context variables used in p_{low}, namely $\{E_f, E_v, d\}$, is a subset of C_{low}. The middleware will invoke the adaptive action *incCurrent* whenever p_{low} is satisfied. The situation $s_{visitor\_nearby}$ is defined similarly in Table 3. Thus, the set of situations $S_{streetlight}$ for the application is $\{s_{low\_illuminance}, s_{visitor\_nearby}\}$. Similarly, the set of context variables used in the application scenario, as listed in Table 1, is given by $C_{streetlight} = \{V_{oc}, I, I_{max}, R, E_v, E_f, d\}$. The set of adaptive actions $U_{streetlight}$ is $\{incCurrent, computIllum\}$. In this way, the CM-centric program is $\langle C_{streetlight}, U_{streetlight}, S_{streetlight}\rangle$. *Cabot* distinguishes a context variable, say I_{max}, from the local variable in an adaptive action by adding a prefix "\$" to the latter, as in $\$I_{max}$.

The system works as follows: When a streetlight detects a visitor within a distance of 5 meters, as specified as the triggering condition in Table 3, the situation $s_{visitor\_nearby}$ is satisfied. The corresponding adaptive action *computIllum* in Table 3 will be invoked to compute the following in sequence: (*i*) the current of the circuit for the given situation based on Ohm's Law, (*ii*) the output power of the streetlight, and (*iii*) the illuminance of incident light at the visitor's location. Similarly, the situation $s_{low\_illuminance}$ will determine whether the actual illuminance satisfies the visitor's

Table 1: Contexts used in sample application

Context	Description
V_{oc}	voltage of open circuit
I	current
I_{max}	maximum current allowed
R	adjustable resistance
E_v	illuminance of incident light at visitor's location
E_f	visitor's favorite level of illuminance
d	distance between visitor and streetlight

Table 2: Constants used in sample application

Constant	Description
ε	maximum tolerance between E_v and E_f
k	physical constant for output power
r	fixed resistance of streetlight
dR	decremental value of adjustable resistance

Table 3: Code fragment for situation *visitor_nearby*

Situation *visitor_nearby* $= \langle C_{in}, p_{in}, computIllum\rangle$	
Contexts	$C_{in} = \{d, V_{oc}, R, I, I_{max}, E_v\}$
Triggering condition	$p_{in} = (d \le 5)$
Adaptive action	*computIllum* $\{$ $\$V_{oc} = V_{oc}$; $\$R = R$; $I = \$V_{oc}/(r+\$R)$; $\$I = I$ $\$P = k*\$I*\$I*r$; $\$d = d$; $E_v = \$P/(\$d*\$d)$; $\}$

Table 4: Code fragment for situation *low_illuminance*

Situation *low_illuminance* $= \langle C_{low}, p_{low}, incCurrent\rangle$	
Contexts	$C_{low} = \{E_f, E_v, d, I, I_{max}, R\}$
Triggering condition	$p_{low} = (E_f - E_v > \varepsilon) \wedge (d \le 5)$
Adaptive action	*incCurrent* $\{$ $\$I = I$; $\$I_{max} = I_{max}$; if $(\$I < \$I_{max})$ & $(R > 0)$ $R = R - dR$; $\}$

favorite level. A satisfaction of $s_{low\_illuminance}$ means insufficient illuminance. It causes the middleware to invoke the adaptive action *incCurrent*, which increases the current by reducing the adjustable resistance. The adjustment will end if the corresponding triggering condition is satisfied. Because of the instability of location sensing [24, 38], however, these context values may further fluctuate. A full description of the application scenario can be found in [21].

3. TESTING CHALLENGES

In this section, we report on our study in identifying three kinds of obstacle that hinder the effective application of standard data flow testing criteria [40] to CM-centric programs because of context-awareness.

Context-Aware Faults: Suppose there is a fault in the triggering condition of the situation *visitor_nearby* which duplicates the triggering condition in the situation *low_illuminance*. (We note that a situation resides at the middleware tier.) In other words, the faulty triggering condition is "$(E_f - E_v > \varepsilon) \wedge (d \le 5)$". Compared with the original triggering condition "$d \le 5$", this fault reduces the chance for the middleware to invoke the adaptive action *computIllum* and, hence, reduces the context-awareness of the application. Traditional data flow testing [12] would compute the data flow associations within the function *computIllum*. Since the fault does not affect these associations, traditional approaches fail to identify the fault effectively. (Our previous work [35] describes another scenario in which traditional data flow testing also fails.) Our present approach takes into account the data flow associations between the adaptive action and the triggering condition. For example, our testing criteria require a test suite to cover the data flow edge between the update of context variable E_v in the last statement of *computIllum* and the use of E_v in the triggering condition. Since the triggering conditions of the situations *visitor_nearby* and *low_illuminance* overlap, such a test case will always invoke the corresponding adaptive actions in pair whenever the situation *visitor_nearby* is satisfied, contradicting the fact that the illuminance may or may not need to increase even when a visitor is in the near proximity. This indicates a failure of the application.

Environmental Interplay: In the smart streetlight example, the context variable d, which denotes the distance of a visitor from a streetlight, is collected via the sensed location of the visitor. The value of d is controlled by the visitor via the middleware, rather than by the application program. Consider the adaptive action *computIllum* again. The statement "$\$d = d$" retrieves the context value from d and assigns it to a local variable $\$d$. Traditional data flow testing would simply require covering a data flow edge from d to $\$d$. It would ignore any potential environmental changes in d. Our criteria require d to be updated environmentally right before the edge is covered in a test execution. During the execution of *computIllum*, suppose d is sensed (wrongly) as a fair large value, say 1000, instead of the correct value of 3. If $\$P$ is small, the value of E_v in *computIllum* will then be changed to a small number, which effectively dims off the light with respect to the visitor. However, the whole purpose of the situation *visitor_nearby* is to serve visitors when they are near the streetlight. Hence, considering environmental updates helps improve the quality of context-aware applications.

Context-Aware Control Flow: Consider a scenario where a visitor walks near a smart streetlight. The favorite level of illuminance E_f and the distance d are continuously sensed by the middleware. At the same time, *computIllum* is continuously invoked to compute the illuminance E_v at the visitor's location. As long as their values satisfy the triggering condition $(E_f - E_v > \varepsilon) \wedge (d \leq 5)$, *incCurrent* will be invoked by the middleware. It is very difficult to enumerate all possible control flow traces for the invocation of an action and to compute the full set of traditional data flow associations. As a tradeoff, our approach captures def-use associations between pairwise invoked adaptive actions. The relevant criterion is defined in Section 5. The evaluation in Section 6 indicates that this approach is still very effective.

4. DATA FLOW ASSOCIATIONS

4.1 Conventional Def-Use Associations

The terminology of our proposed data flow criteria closely resembles that of conventional approaches in [12, 15, 27]. We shall first summarize the latter in this section. We model a standard program in an application as a control flow graph (CFG) $G = (N, E)$, where N is a set of nodes and E is a set of edges. We assume that every CFG starts with a unique entry node. A *complete path* is a path in the CFG starting from the entry node and ending at an exit node [12]. The storage of the value to a variable v, such as the occurrence of v on the left-hand side of an assignment statement, is a *definition* (or *def*) of v. A *use* or *usage* of a variable v refers to the fetching of a value of v, such as an occurrence of v on the right-hand side of an assignment statement or in a predicate of a statement [15].

A sub-path $\langle n_i, \ldots, n_j \rangle$ in a CFG is said to be *definition-clear* with respect to the variable x when none of $n_i, \ldots, n_j$ defines or undefines x [12]. A *def-use association* is defined as a triple $\langle x, n_d, n_u \rangle$ such that the variable x is defined at node n_d and used at node n_u, and there is a definition-clear sub-path with respect to x from n_d to n_u, exclusively [12]. For the ease of presentation, we define a predicate $def\_clear(\langle x, n_d, n_u \rangle)$ to be true whenever $\langle x, n_d, n_u \rangle$ is a def-use association.

A complete path π_x is said to *cover* a def-use association $\langle x, n_d, n_u \rangle$ if a sub-path of π_x from node n_d to node n_u, exclusively, is a definition-clear path with respect to x [12]. Thus, if any of the definitions can reach a use in a CFG from the entry node of CFG via a definition-clear path, a def-use association is identified. The set of all *reaching definitions* at node n in CFG_i is denoted by $RD_i(n)$ [13, 14].

4.2 Context-Aware Definitions and Uses

For a standard program, the conventional def-use associations presented in Section 4.1 aim at relating a definition of a variable to a use of the variable via a definition-clear sub-path. A CM-centric program extends a standard program with a set of context variables and a set of situations. The data flow associations need to be extended accordingly.

A *context variable* is a special type of variable. It serves as an ordinary variable in a standard program and, at the same time, is used for exchanging contexts with the pervasive environment. In this section, we examine the definitions and usages of variables and, in particular, context variables.

4.2.1 Definitions of Variables

According to existing proposals of context-aware systems [37, 39], a context variable can be defined and updated via either an *assignment statement* or *sensing the environmental contexts*. They are described as follows:

Type 1: An **update via an assignment statement** refers to the occurrence of a variable in a statement that stores the value of the variable. It coincides with the definition of a variable in the conventional def-use association [12]. Consider the statement "$n12 : E_v = \$P/(\$d * \$d)$" of the program unit *computIllum* in Figure 1, where E_v is a context variable. The occurrence of E_v is said to be a definition of the context variable E_v.

Type 2: An **environmental update**, or an update by sensing environmental contexts, is achieved through the environmental context acquisition module of the context-aware middleware. For example, the voltage of the open circuit, V_{oc}, is read from the sensor of the smart streetlight device. It is analogous to updating the content of a memory location (or a variable) of a program through an external means. It is not attained by an explicit programming construct in a program unit, such as an assignment operator. The "definition" of a variable in conventional def-use association cannot capture this type of update and should be extended.

DEFINITION 4 (DEFINITIONS (DEF) OF VARIABLES). *A* **definition** *or* **def** *of a variable v is an occurrence of v (i) in a statement where a value is assigned to v, or (ii) in a statement where v is a context variable that can be instantiated by the sensing environmental contexts.*

For simplicity, we refer to a definition of a context variable c as a *context definition*, denoted by def_c. We further use Def_c to represent the set of all the context definitions of the context variable c in the program.

4.2.2 Usages of Variables

We continue to examine the usages of variables in a CM-centric program. For ordinary variables, the conventional definition of usage of a variable strictly applies. We do not observe any difference. For context variables, a usage occurrence of a context variable can occur in either a program unit or a situation. We call them action use and situation use, respectively. They correspond to the only two types of position for developers to code the reference of a variable in a CM-centric program. They are further elaborated as follows:

An *action use* coincides with the conventional definition of a use of a variable. It refers to an occurrence of a context variable in a statement that fetches the value of the variable, such as on the right-hand side of an assignment statement.

A *situation use* is the usage of a context variable in the triggering condition of a situation. For instance, in the situation $s_{low\_illuminance}$ in Table 4, all the occurrences of the context variables E_f, E_v, and d are situation use. Note that a situation use can be recorded only if the corresponding situation is satisfied; otherwise the usage is not observable.

DEFINITION 5 (USAGES (OR USES) OF VARIABLES). *A* **usage** *or* **use** *of a variable v is an occurrence of v (i) in a statement that fetches the value of v (known as* **action use***), or (ii) in the triggering condition of a situation if v is a context variable (known as* **situation use***).*

We denote a situation use of a context variable c by $suse_c$ and an action use by $ause_c$. The set of uses, the set of situation uses, and the set of action uses of c are denoted by Use_c, $SUse_c$, and $AUse_c$, respectively.

4.2.3 Update-Uses of Variables

It is apparent from Definitions 4 and 5 that a context definition and a usage of the same context variable may occur in the same statement. Consider, for instance, the context variable d in the code in Table 3. Suppose the current value of d sensed by the middleware is d_0, where $d_0 \leq 5$. Thus, the triggering condition is satisfied and *computIllum* is invoked. Let us focus on the statement $\$d = d$. Suppose there is an environmental update of the variable d immediately before the execution of this statement. The value of the variable d in the statement no longer depends on the previous value d_0. Instead, it will depend on the environmental update.

Such a scenario of an environmental update followed by a usage occurrence of the same variable is possible if a context variable appears in the triggering condition of a situation, in a statement that fetches the value of the variable, or in the predicate of a statement. As a result, an occurrence of an environmental update can be determined from the source code by determining a usage occurrence of a context variable. We refer to such an occurrence of the environmental update as an update-use occurrence and define it as follows:

DEFINITION 6. (UPDATE-USE OCCURRENCES OF VARIABLES). *An* **update-use** *occurrence of a context variable c is an occurrence containing a context definition def_c and a context use use_c of c, where def_c refers to the instantiation of c due to the sensing of environmental contexts.*

For simplicity, the set of context definitions of a context variable c involving update-use occurrences will be denoted by $UDef_c$. The set of other context definitions of c will be denoted by $ODef_c$. Thus, $Def_c = UDef_c \cup ODef_c$ while $UDef_c \cap ODef_c = \emptyset$.

Based on the above observations, we define a *Context-aware Flow Graph* (*CaFG*) to describe the control and data flow information in a CM-centric program. Owing to space limitation, only part of the CaFG of the smart streetlight example is shown in Figure 1. The procedure for constructing a CaFG can be found in [21]. We briefly describe a CaFG as follows:

Each situation is modeled as a CFG known as a *situation graph*. It consists of the entry node, an exit node, a predicate node representing the triggering condition, and a node that follows the true branch of the predicate node and invokes the adaptive action. For each node in any CFG that contains at least one update-use occurrence, we transform it into an *update-use node*, which leads to a subgraph called an *update-use subgraph*.

The update-use subgraph consists of the entry node, an exit node, nodes that we shall call *use nodes*, and phantom nodes. A sample update-use subgraph for the update-use node $n2$ of Figure 1 is shown in Figure 2. A use node, depicted as a standard rectangle, is annotated with the original statement of the update-use node but does not represent an update-use occurrence. For each update-use occurrence, a *phantom node* representing an environmental update is introduced. A control flow edge, which we call a *phantom control flow*, will connect the entry node via a phantom node to the corresponding use node. When entering the subgraph, one may choose whether or not to visit a phantom node, that is, whether or not to have an environment update.

Figure 1 shows the CaFG of the streetlight application program. Situation graphs are created for $s_{low\_illuminance}$ and $s_{visitor\_nearby}$. Update-use subgraphs can be obtained from update-use nodes $s1$, $s4$, $n1$, $n2$, $n3$, $n4$, $n6$, $n7$, $n9$, and $n11$ involving the context variables E_f, E_v, V_{oc}, d, R, I, and I_{max}.[5]

4.3 Def-Use Associations for CM-Centric Programs

We are interested in determining the definition-usage relationships in a CM-centric program with a view to conducting data flow testing. In this section, we discuss the data flow associations that we find to be special to context variables in CM-centric programs. For ordinary variables, we adopt the conventional approach to compute data flow associations (see Section 4.1).

Variables in a CM-centric program can appear in program units, in situations, or both. As it is infeasible to have an assignment statement for a situation in a CM-centric program, there are, by exhaustion, three types of def-use association connecting a definition and a usage of any variable. Type (DU1): Both a definition occurrence and a usage occurrence of a variable appear in a situation. Type (DU2): A definition occurrence of a variable appears in a program unit and a usage occurrence appears in a situation. Type (DU3): Both occurrences are within program units. We elaborate on each of these types below.

We recall that the only way to update a context variable in a situation is through an environmental update, and the only way to specify a reference of a context variable in a situation is via a usage occurrence of the variable. Hence, type (DU1) refers to an update-use of a context variable c in a node n_s followed by a situation use of c in a node n_e such that there is a definition-clear path with respect to c from n_s to n_d. We note that every context variable in a CM-centric program is unique. An environmental update of c will affect all occurrences of this variable, which should include its occurrence in node n_e. The only possible definition-clear path is, therefore, $\langle n_e, n_e \rangle$. Thus, the definition-usage association is

[5] Only the update-use subgraph of $n2$ is shown in Figure 2 because of space limitation.

Figure 1: Context-aware flow graph for streetlight program

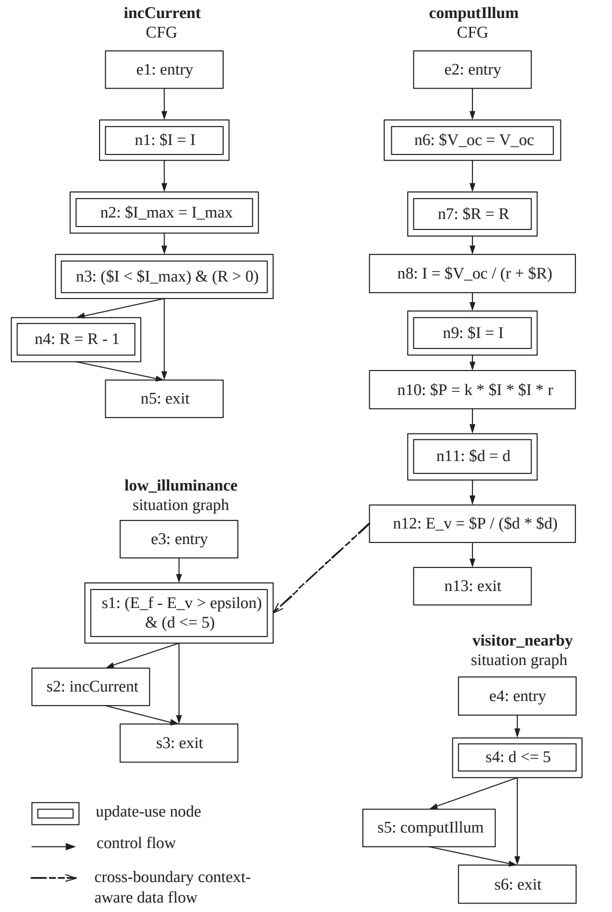
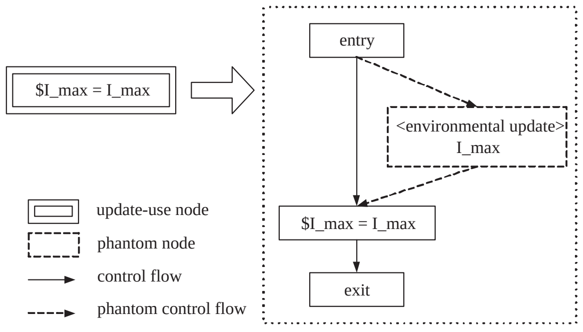


Figure 2: Update-use subgraph of node $n2$ in Figure 1



$\langle c, n_e, n_e \rangle$, where n_e is a triggering condition of a situation. In Figure 1, for example, the def-use associations of type (DU1) include $\langle E_f, s1, s1 \rangle$, $\langle E_v, s1, s1 \rangle$, and $\langle d, s1, s1 \rangle$.

Type (DU2) refers to the case when a definition occurrence of a variable appears in a program unit or situation while a usage occurrence of the same variable appears in a situation. Based on this intuition, we formalize it as a def-situ association as follows:

DEFINITION 7 (DEF-SITU ASSOCIATIONS). *A* **def-situ association** α *for a context variable c is a triple $\langle c, def_c, suse_c \rangle$ when there is a situation $\langle C, p, Act \rangle$ such that (i) $suse_c$ is a situation use at p, (ii) $c \in C$, (iii) $p(C) = true$, and (iv) $def\_clear(\langle c, def_c, suse_c \rangle) = true$.*

Figure 3: The def-situ associations algorithm

```
Algorithm  COMPUTE DEF-SITU
  Input    The CaFG for a CM-centric program $\langle C, U, S \rangle$;
  Output   The set DEF-SITU of all def-situ associations;
  Associate every node in the CaFG of every program unit with the
     set of context definitions;
  Associate every node in each situation graph with the set of
     update-use occurrences and situation uses;
  for each $c \in C$
    Collect $UDef_c$, $ODef_c$, and $SUse_c$;
    if $ODef_c = \emptyset$
      for each $def_c \in UDef_c$ and each $suse_c \in SUse_c$
        if $def\_clear(\langle c, def_c, suse_c \rangle) = true$
          Add $\langle c, def_c, suse_c \rangle$ to DEF-SITU;
    else
      for each $def_c \in ODef_c$ and each $suse_c \in SUse_c$
        if $def\_clear(\langle c, def_c, suse_c \rangle) = true$
          Add $\langle c, def_c, suse_c \rangle$ to DEF-SITU;
```

Figure 1 highlights a def-situ association $\langle E_v, n12, s1 \rangle$, annotated as a *cross-boundary context-aware data flow*. We note that type (DU2) is a (deliberate) generalization of type (DU1).

The algorithm to compute the set of def-situ associations for a CM-centric program $\langle C, U, S \rangle$ is given in Figure 3. Suppose that the number of lexical tokens of the program is n and each evaluation of def_clear takes time m. The algorithm will take $O(|C|mnlog(n))$ time [21].

Type (DU3) corresponds to the circumstance where both the definition and usage in a def-use association occur within program unit(s). It covers the case where the definition and the usage appear in the same program unit as well as the case where they appear in different units. This type is built on top of Definition 7. We recall that a def-situ association α defines a data flow edge from a context definition of a variable c in action Act_i to a situation use of the variable in a situation. The association requires the corresponding situation to be satisfied. Once the situation is satisfied, the associated adaptive action Act_j should be invoked. We use the notation $Act_i :\overset{\alpha}{\Rightarrow} Act_j$ to indicate that Act_i carries out the context definition of the triple α and Act_j is invoked by the situation that receives the situation use of α. We shall first present our definition and then show the algorithm to compute these associations. Since it is generally undecidable to determine whether a path is feasible, we compute, instead, the related data flow associations between two consecutive calls of adaptive actions in a CM-centric program.

DEFINITION 8 (PAIRWISE DU ASSOCIATIONS). *A* **pairwise context-aware def-use association** *(or simply a* **pairwise du association***) du with respect to (i) two program units Act_i and Act_j and (ii) a def-situ association $\alpha = \langle c, def_c, suse_c \rangle$ such that the relationship $Act_i :\overset{\alpha}{\Rightarrow} Act_j$ holds, is of one of the following two subtypes:*

(a) $du = \langle c', def_{c'}, ause_{c'} \rangle$ in which $def_{c'}$ is from Act_i, $ause_{c'}$ is from Act_j, $def_{c'} \in RD_i(def_c)$, and $def\_clear(c', e_j, ause_{c'}) = true$, where e_j refers to the entry node of Act_j.

(b) $du = \langle x, def_x, use_x \rangle$ in which both def_x and use_x occur in Act_j, and $def_x \in RD_j(use_x)$.

In Figure 1, $\langle I, n8, n1 \rangle$, where I is a definition at $n12$ that can reach $n1$, is an example of subtype (a) according to Definition 8. It is derived from the relationship $visitor\_nearby :\overset{\alpha_1}{\Rightarrow} low\_illuminance$, in which $\alpha_1 = \langle E_v, n12, s1 \rangle$. Similarly, still for the relationship

Figure 4: The pairwise du associations algorithm

```
Algorithm   COMPUTE PAIRWISE-DU
  Input     A def-situ association $\alpha = \langle c, def_c, suse_c \rangle$ such that
                 $Act_i \overset{\alpha}{:\Rightarrow} Act_j$;
  Input     $G_i$ and $G_j$, respective CFGs for $Act_i$ and $Act_j$;
  Output    The set $PAIRWISE\text{-}DU(\alpha)$ of all pairwise du
                 associations with respect to $\alpha$;
  $RD_i(def_c)$ = the set of reaching definitions at $def_c$ in $G_i$;
  for each variable $c'$ such that there exists $def_{c'} \in RD_i(def_c)$
     for each node $nd$ in $G_j$
        if $nd$ has a use of $c'$ and $def\_clear(c', e_j, nd) = true$
        // Note: $e_j$ is the entry node of $Act_j$
           Add $\langle c', def_{c'}, nd \rangle$ to $PAIRWISE\text{-}DU(\alpha)$;
  for each node $nd$ in $G_j$
     $RD_j(nd)$ = the set of reaching definitions at $nd$ in $G_j$;
     for each variable $x$ that has a use at $nd$
        if $x$ has a definition $def_x$ in $G_j$ such that $def_x \in RD_j(nd)$
           Add $\langle x, def_x, nd \rangle$ to $PAIRWISE\text{-}DU(\alpha)$;
```

$visitor\_nearby \overset{\alpha_1}{:\Rightarrow} low\_illuminance$, the pairwise du associations of subtype (*b*) include $\langle \$I, n1, (n3, n4) \rangle$ and $\langle \$I_{max}, n2, (n3, n4) \rangle$,[6] to name a couple.

The algorithm for pairwise du associations is shown in Figure 4. The time cost is $O(|C|mn^2)$ [21], where $|C|$ is the total number of context variables in the CM-centric program, n is the total number of lexical tokens, and m is the time cost in evaluating a *def_clear* once.

5. TEST ADEQUACY CRITERIA

To measure the quality of a test set for a CM-centric program, we define in this section our test adequacy criteria using the data flow associations described in Section 4. First of all, we recall that a test adequacy criterion C_1 is said to *subsume* a test adequacy criterion C_2 if every test set that satisfies C_1 will also satisfy C_2 [40].

According to Definition 3, a CM-centric program differs from a standard program by having a set of context variables and a set of situations. Our first test adequacy criterion is to exercise every situation at least once. Whenever a situation is satisfied, the middleware will invoke the corresponding adaptive action. On the other hand, an adaptive action may be invoked by another adaptive action. Hence, a test set that covers all adaptive actions does not necessarily cover all situations.

CRITERION 1 (ALL SITUATIONS). *A test set T satisfies the* **all-situations criterion** *for a CM-centric program $\langle C, U, S \rangle$ if and only if, for each situation $s = \langle C_s, p, Act \rangle$ in S, the complete path of at least one test case $t \in T$ includes the predicate node of s such that $p(C_s) = true$.*

We refine the all-situations criterion to cover all def-situ associations, whose definition is as follows:

CRITERION 2 (ALL DEF-SITU ASSOCIATIONS). *A test set T satisfies the* **all-def-situ-associations criterion** *for a CM-centric program $\langle C, U, S \rangle$ if and only if, for every def-situ association α obtained from the algorithm COMPUTE DEF-SITU, there is at least one test case $t \in T$ such that $def\_clear(\alpha) = true$.*

We note that a usage occurrence in a situation should either be an update-use or have a corresponding context definition in a program unit. Moreover, according to condition (*iii*) of Definition 7, the corresponding situation should be satisfied. Hence, a test set

[6] We adopt the standard representation [12, 40] for predicate uses inside a program unit.

Figure 5: Subsumption hierarchy of context-aware test adequacy criteria

$$\begin{tikzcd}
\text{all-pairwise-du-associations} \arrow[d] \\
\text{all-def-situ-associations} \arrow[d] \\
\text{all-situations}
\end{tikzcd}$$

satisfying the all-def-situ-associations criterion should also satisfy the all-situations criterion. In this connection, the all-def-situ-associations criterion subsumes the all-situations criterion.

We further refine the above criterion by linking up a definition occurrence and a usage occurrence in two adaptive actions through a situation. If the above criterion is seen to address type (DU2), then we want our third criterion to address type (DU3). In brief, for every def-situ association α, this criterion requires an adequate test set to exercise all pairwise du associations that are relevant to α. It is simple to observe from condition (*i*) of the following definition that the criterion subsumes the all-def-situ-associations criterion.

CRITERION 3 (ALL PAIRWISE DU ASSOCIATIONS). *A test set T satisfies the* **all-pairwise-du-associations criterion** *for a CM-centric program $\langle C, U, S \rangle$ if and only if the following two conditions are satisfied: (i) T satisfies the all-def-situ-associations criterion; and (ii) for every pairwise du association du obtained from the algorithm COMPUTE PAIRWISE-DU(α), there is at least one test case $t \in T$ such that def_clear(du) = true.*

We note that the definition of α can be generalized to be a series of adaptive actions. By so doing, more general form of the above two criteria can be developed. The extension is not difficult.

Figure 5 shows the subsumption hierarchy for the family of test adequacy criteria defined in this section.

6. EVALUATION

This section reports on the experimentation results of our test adequacy criteria, namely, all-situations, all-def-situ-associations and all-pairwise-du-associations. We also benchmark our proposal against Frankl and Weyuker's all-uses criterion [12]. Furthermore, the study is also a novel attempt to test RFID-based systems. Our prototype testing tool automates the test set generation and evaluation processes.

6.1 Experimental Design

We use *Cabot* [37] and its evaluation application as our testbed. The testbed consists of a middleware supporting context-aware reasoning and triggering, and an application implementing the *LANDMARC* RFID-based location sensing algorithm [24].

Location-sensing applications are popular as pervasive computing applications. RFID contexts are particularly important because they carry vital environmental information that the applications need to handle. As we have introduced in Section 1, they are recognized as an important enabling technology for supply chain management. However, like other typical environmental information such as room temperature and air pollution level, RFID contexts are captured via sensors and are intrinsically imprecise and approximate. Applications using RFID contexts, such as location-sensing programs, should contain program logic to tackle such kinds of approximation, apart from having business logic that can be developed independently of context-aware computing.

Cabot continually collects radio frequency strength signals (context values) from various RFID tags. Based on the different radio frequency strength signals and other environmental contexts, the middleware triggers different adaptive actions to allow the application to estimate the location of a tracking tag.

Our experiment focuses on the testing of the *LANDMARC* application. We use the original implementation as the golden version to check for failures of the faulty versions. The experiment consists of three steps.

Firstly, our tool generates 40 test sets for each adequacy criterion given in Section 5. For comparison purpose, we also generate 40 test sets based on the standard all-uses criterion [12]. The test cases are randomly selected from a test pool containing 8 000 radio frequency strength signal data collected from different locations of tracking tags from online RFID sensor networks. Since full coverage of testing criteria is often infeasible in practice, an upper bound on the number of trials in selecting test cases is set for each criterion. A test set generation process thus stops when full coverage is achieved or the upper bound has been reached. A test case that increases the coverage of the criterion concerned will be added to the adequacy test set. It also reports the percentage of coverage. Our automated method to select test cases is similar to that described in [15].

Figure 6 shows the algorithm that our tool uses to generate test sets for the all-pairwise-du-associations coverage. The algorithm accepts the test pool and a hashtable *dua* as inputs and produces an adequacy test set as output. *dua* is a hierarchical hashtable having an entry for each def-situ association, as well as an entry for each pairwise du association related to the respective def-situ associations. For a sequence of test cases, *dua* is covered in a cumulative way; the test cases will achieve full coverage when all the pairwise du associations in *dua* are marked as covered.

In the algorithm, *trace* is a sequence of variable definition and usage occurrences that are recorded when a test case is selected and forced to execute. For any individual pairwise du association covered, the algorithm carries out two rounds of forward and backward searches to find out the corresponding definition-clear path.[7]

It is worth noting that the construction of test sets for all-situations and all-def-situ-associations criteria are byproducts of the algorithm EVALUATE *PAIRWISE-DU*.

Secondly, we create 50 faulty versions, each having one fault. There are 18, 11, 13, and 8 faults in the context repairing logic, context-aware rules, situational application configuration, and estimation logic, respectively. Simple faults such as data conversion format mismatches are excluded. All versions are assured by an experienced software developer.

Finally, our tool applies all the generated test sets to all versions to evaluate their fault-detection capabilities. For each adequacy criterion, the *fault-detection rate* [11] is defined as the ratio of the number of corresponding adequacy test sets (each containing at least one test case that exposes the fault) to the total number of corresponding adequacy test sets with respect to the criterion. For the purpose of comparison, we use all 8 000 test cases from the test pool to test every faulty version. The corresponding percentage of failed test cases denotes the failure rate of the version. Our tool also compares automatically the execution results of faulty versions and that of the golden version. It computes the fault-detection rate for each criterion.

[7] Suppose the average length of an execution trace is L and the number of context variables is $|C|$. Then, for each test case t selected, the time cost to evaluate the coverage of t will be $O(|C|L^4)$. When the upper bound of the number of selected test cases is N, the total time cost to generate a test set for the all-pairwise-du-associations criterion will be $O(N \cdot |C|L^4)$.

Figure 6: Algorithm for generating all-pairwise-du-associations adequacy test set

```
Algorithm   EVALUATE PAIRWISE-DU
  Input     TestPool;
  Input     dua, the set of pairwise du associations;
  Output    TestSet for the all-pairwise-du-associations criterion;
  int $i = 0$;
  while dua is not fully covered and $i < UPPER\_BOUND$
      $i++$;
      Randomly select a non-redundant $t$ from TestPool;
      Execute $t$ to obtain trace;
      for each element $e_{f_1}$ in forward search of trace
          if $e_{f_1}$ is a situation use of context variable $c$
              for each element $e_{b_1}$ in backward search from $e_{f_1}$
                  if $e_{b_1}$ is a definition of $c$
                      if dua contains an entry $\langle c, e_{b_1}, e_{f_1} \rangle$
                          subdua = the set of pairwise du
                              associations of $\langle c, e_{b_1}, e_{f_1} \rangle$;
                          SEARCH-DU($e_{f_1}$, subdua);
      if the coverage of dua increases
          Add $t$ to TestSet;
Procedure   SEARCH-DU($e_{f_1}$, subdua)
  for each element $e_{f_2}$ in forward search from $e_{f_1}$
      if $e_{f_2}$ is a usage occurrence of variable $v$
          for each element $e_{b_2}$ in backward search from $e_{f_2}$
              if $e_{b_1}$ is a definition of $c$
                  if subdua contains an entry $\langle v, e_{b_2}, e_{f_2} \rangle$
                      Mark $\langle v, e_{b_2}, e_{f_2} \rangle$ in dua as covered;
```

Table 5: Statistics of test sets generated according to different adequacy criteria

Criterion	Coverage (in %) of generated test sets (40 sets per criterion)		
	min.	avg.	max.
all-situations	96.3	99.4	100
all-def-situ-associations	93.8	96.6	99.2
all-pairwise-du-associations	86.7	87.1	87.4
all-uses	71.2	72.6	72.9

6.2 Data Analysis

Through the analysis by Algorithms COMPUTE *DEF-SITU* and COMPUTE *PAIRWISE-DU*, the *LANDMARC* testbed program has a total of 81 situations, 129 def-situ associations, 7 682 pairwise du associations, and 177 def-use associations with reference to the standard all-uses criterion. Our tool generates 40 test sets for every test adequacy criterion. For the all-situations, all-def-situ-associations, all-pairwise-du-associations, and all-uses adequacy criteria, we set the upper bound of number of selected test cases to be 300, 1 000, 2 000, and 2 000. The generated test sets are summarized in Table 5.

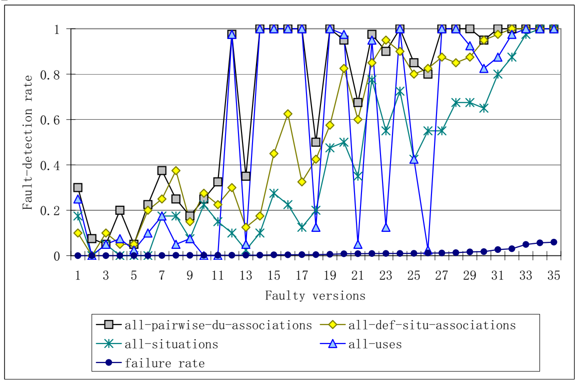
After applying the entire test pool to all the 50 faulty versions, we find 8 faults which cannot be exposed by any test case. In addition, there are 6 faulty versions that have failure rates higher than 0.95 and another version with a failure rate of 0.34. A deeper investigation of these 7 highly detectable faults shows that all of them are located on the key paths of the module to compute the final estimated tracking location, so that they have high chances to be exposed. Since these 15 faulty versions may not be suitable for determining the fault-detection capabilities of the adequacy test sets, they are discarded in the sequel. The remaining 35 faulty versions have failure rates within a range of 0.0003 to 0.06, with an average of 0.011.

We first compare the criteria irrespective of the failure rates of faulty versions. The overall fault-detection rates of the three adequacy criteria are given in Table 6. All the faulty versions are further categorized according to whether a fault is related

Table 6: Overall fault-detection rates

Type of faulty versions	Fault-detection rate of adequacy test sets			
	Criterion	min.	avg.	max.
Situations (10 faulty versions)	all-situations	0.075	**0.548**	0.875
	all-def-situ-associations	0.15	**0.723**	1.0
	all-pairwise-du-associations	0.175	**0.878**	1.0
	all-uses	0.025	**0.730**	1.0
Actions (25 faulty versions)	all-situations	0.0	**0.326**	1.0
	all-def-situ-associations	0.0	**0.496**	1.0
	all-pairwise-du-associations	0.05	**0.617**	1.0
	all-uses	0.0	**0.472**	1.0
Total (35 faulty versions)	all-situations	0.0	**0.389**	1.0
	all-def-situ-associations	0.0	**0.544**	1.0
	all-pairwise-du-associations	0.05	**0.691**	1.0
	all-uses	0.0	**0.546**	1.0

Figure 7: Fault-detection rates of adequacy criteria with respect to failure rates



to a situation or an action. Out of the 35 faulty versions, 10 and 25 belong to the respective category. For each category, the minimum, average, and maximum fault-detection rates with respect to each criterion are computed. On average, the all-pairwise-du-associations criterion gives the highest fault-detection rates.

In order to study the change of fault-detection rates with respect to failure rates, we present line plots of the fault-detection rates of various adequacy criteria against the faulty versions, as shown in Figure 7. The 35 faulty versions are sorted along the horizontal axis in ascending order of their failure rates. The vertical axis scales the values of the fault-detection rates. The values for the same adequacy criterion are connected to demonstrate the trend of fault-detection with the increase of failure rates. We find that the fault-detection rates of the three criteria proposed in Section 5 generally increase with the increase of failure rates. On the other hand, the fault-detection rate of the all-uses criterion fluctuates greatly, ranging from less than 0.1 to nearly 1.0, along the middle segment of the failure rate.

A finer inspection of Figure 7 shows that the greatly fluctuating segment of the all-uses line starts from the 12th faulty version and ends at the 26th faulty version, representing failure rates from 0.002 to 0.012. We therefore classify the 35 faulty versions into three intervals known as A, B, and C, partitioned by the failure rates of 0.002 and 0.012. Figure 8 shows the fault-detection rates of the four criteria in each of the intervals. We find that all the criteria are promising in interval C. In interval B, only all-pairwise-du-associations have a detection rate of more than 0.8, outperforming all-uses by 37%. In interval A, the fault-detection rate of every criterion is below 0.2 as a result of the very low failure rate (average 0.0007). In particular, the effectiveness of all-uses drops most rapidly in interval A to become the lowest among all criteria.

Figure 8: Fault detection rates with respect to different failure rate intervals

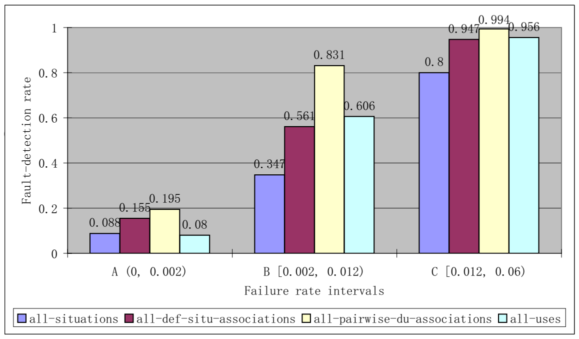


Table 7: Effectiveness comparison of different criteria

	Description	Number of faulty versions			
		Total	Failure rate intervals		
			(A)	(B)	(C)
1	$C_{pdu} > C_u$	15	5	7	3
2	$C_{pdu} = C_u$	20	5	9	6
3	$C_{ds} > C_u$	10	2	6	2
4	$C_u > C_{ds}$	11	1	8	2
5	$(C_{pdu} = C_u) \wedge (C_{pdu} = C_{ds}) \wedge (C_{ds} = C_u)$	8	4	0	4
6	$(C_{pdu} > C_{ds}) \wedge (C_{pdu} > C_s)$	15	2	10	3

For a better comparison of fault-detection effectiveness of different criteria, we carry out statistical hypothesis testing [31] based on the collected data. The method is similar to that in [11]. For any two criteria C_1 and C_2 under comparison, let r_1 and r_2 be their respective fault-detection rates. In the following, we set the null hypothesis as $H_0 : r_1 \leq r_2$ and the alternate hypothesis as $H_1 : r_1 > r_2$ with a significance level of 0.05. We use the notation $C_1 > C_2$ to denote the result that the null hypothesis $r_1 \leq r_2$ can be rejected, meaning that r_1 is significantly higher than r_2. In addition, we also use the notation $C_1 = C_2$ to represent the result that another null hypothesis of neither $r_1 \leq r_2$ nor $r_2 \leq r_1$ can be rejected, meaning a comparable effectiveness of C_1 and C_2 in detecting a certain fault.

Table 7 compares the effectiveness of different categories of criteria. The all-situations, all-def-situ-associations, all-pairwise-du-associations, and all-uses criteria are denoted by C_s, C_{ds}, C_{pdu}, and C_u, respectively. The "Description" column lists the hypothesis tests that the category of faulty versions have passed. For instance, the first row illustrates that the category $C_{pdu} > C_u$ contains a total of 15 faulty versions, in which the fault-detection rate of all-def-situ-associations is significantly higher than that of all-uses. Moreover, the 15 faulty versions are classified into the three failure rate intervals shown in Figure 8. It is worth noting that any two of the categories given in Table 7 are not necessarily mutually exclusive.

7. DISCUSSIONS

Our preliminary experimentation shows that our adequacy criteria are promising in measuring the quality of test sets for CM-centric programs. Firstly, our technique has been demonstrated to effectively handle faults related to contexts. As shown in Table 6, the all-pairwise-du-associations criterion has an average fault-detection rate of 0.878 for detecting faults related to situations.

Secondly, among our three adequacy criteria, the all-pairwise-du-associations criterion demonstrates a fault-detection rate which is higher than or equal to the standard all-uses criterion [12]. The hypothesis test in Table 7 supports that the former is statistically more effective than the latter in detecting 15 out of all the 35 faults, while the latter never outperforms the former. On the other hand, the all-def-situ-associations and all-uses criteria have comparable fault-detection rates in general (see Table 6).

The fluctuating behavior of the all-uses criterion shown in Figure 7 is interesting. In interval B, the fault-detection rate of the all-uses criterion is by no means steady. We have computed the standard deviations of the fault-detection rates for all the data flow criteria reported. The one for the all-uses criterion is 0.443, which is almost twice as that of any of the other three criteria. This phenomenon warrants further investigation.

We would also like to highlight potential threats to validity of the experiment. For internal validity, we note that the experiment uses one program and fault types relevant to the program to evaluate our criteria. Its test pool is based on 8 000 real RFID data only. The 8 faults that could not be exposed by any test case might actually be detected by further testing if a larger test pool were available. In addition, our prototype testing tool may contain faults, even though it has been "thoroughly" tested. More empirical studies are warranted to further evaluate the proposal.

8. RELATED WORK

In this section, we review related literature on the testing of context-aware software systems. To do this, we first briefly review the background of *Cabot* [37, 38], the middleware on which our evaluation has been conducted.

Context-aware computing is an important research topic of pervasive computing [30]. According to the survey in [6], the *Active Badge* system [36] is the pioneering study. *Context Toolkit* [7] is one of the earliest frameworks for developing generic context-aware applications. It provides abstract components for programmers to build context-aware systems and obtain context data from sensors conveniently. However, the framework is centralized, low-level, and without middleware support.

Emmerich [8] and Roman et al. [29] suggest that the middleware-based approach is a trend in software engineering to handle the mobile and pervasive environment. During the last few years, there have been many representative proposals in context-aware pervasive computing using a middleware-based architecture [3, 4, 5, 17, 28, 39]. Among these, *RCSM* [39] is the first general-purpose context-aware middleware proposal that supports context-aware behaviors in smart devices. Various other models with reflective context-aware middleware, such as *CARMEN* [3], *CARISMA* [4], and *MobiPADS* [5], have been developed to support reflective reasoning and resource management in response to context changes. *Gaia* [28] is a middleware-based framework supporting the development and management of smart spaces in the pervasive environment. It incorporates an infrastructure for context-awareness based on first order logic [26]. While its middleware-based context-aware rule system and context-triggered applications can be a suitable testbed for our techniques, whether our testing criteria can be effectively applied to rules in first order logic requires further empirical studies. *EgoSpaces* [17] is a middleware that supports context-aware programming. It is built on the mobile tuple space system *LIME* [23]. Contexts are represented as tuples and captured via patterns while programs are activated upon matching of constraints, which is similar to the mechanisms in *Cabot*. In short, our testing notions and techniques can be adapted to quite a number of representative context-aware platforms with some adaptations such as interfacing requirements. Since *Cabot* is designed from the mindset of quality assurance, it is a good platform for us to evaluate our testing proposal.

The prototype application of *Cabot* implements the RFID-based location sensing algorithm *LANDMARC* [24] and uses the middleware to context-sensitively adjust detected context information so as to reduce the average errors of the sensed locations. In the present work, *Cabot* with its *LANDMARC* implementation is used as the testbed. Details of the empirical study have been described in Section 6.

In the rest of this section, we review the testing research for pervasive computing. To our best knowledge, it is not plentiful. Axelsen et al. [2] propose a specification-based approach to test reflective software in an open environment, and use a random selection strategy to produce test inputs. When the execution sequence of any test input violates the specifications, it detects a failure. Flores et al. [9] apply temporal logic to define contextual expressions in context-aware software. They then use a kind of category partitioning to break up the temporal logic expressions. Intuitively, they require each partition to be covered. However, their work does not provide test case generation guidelines. Our previous work [35] basically generates multiple context tuples as test cases to check whether the outcomes satisfy isotropic properties of context relations. It is a black-box approach with a focus on the test oracle problem. The proposal presented in this paper is a program-based approach that does not rely on any specifications. It focuses on the test set adequacy problem.

We also review constraint-based approaches as the context-aware interface can be regarded as a set of constraints. Tai [33], among others such as [18, 19, 34], investigate a rule-based approach to construct predicate-based test cases. He focuses on specifications of programs consisting only of predicate rules. These rules cannot trigger adaptive actions that may produce instant side effects to other parts of the same expression. Jin and Offutt [16] propose a set of constraint-based test adequacy criteria for software systems that need to interact with the environment. However, their target programs have explicit "system calls" to identify the locations where the environment can affect the program or be affected.

When compared with researches in test adequacy criteria of standard programs, our notions are in line with them [12, 27]. Various empirical studies in conventional data flow testing are performed and evaluated in [10, 11, 15]. Investigations on data flow testing at the integration level can be found in [13, 25]. The data flow approach has been applied to the testing of other non-conventional programs as well. Souter and Pollock [32] investigate different levels of def-use associations for objects with different levels of calling paths, and apply them to increase the test coverage of object-oriented programs. Kapfhammer and Soffa [20] propose a family of test adequacy criteria based on the hierarchy of relational database entities and a novel notion of database interaction graph to capture the def-use associations for database entities. Our approach is in line with theirs. We capture def-use associations among entities relevant to pervasive contexts.

9. CONCLUSION

Pervasive context-aware software is a novel kind of applications that challenges existing testing techniques. In this paper, we have investigated the testing of context-aware middleware-centric pervasive applications. We have formalized the notion of context-aware data flow entities. Based on them, we have proposed a novel family of test adequacy criteria to measure the quality of test sets. Corresponding algorithms are given. We have also reported the experimentation results of an RFID-based location-sensing system, in which we applied a testing tool to automate the construction of adequate test sets and the evaluation of test results. The empirical results are promising.

Our approach is applicable and effective to pervasive computing. The context-aware program logic normally spans over the application tier and the middleware tier. The part residing in the middleware tier is ignored by conventional testing methods. Failures specific to context-aware features relevant to the middleware is, therefore, difficult to be exposed by conventional testing techniques.

We plan to conduct more case studies on different applications on other platforms to further evaluate our approach. For programs with very low failure rates (< 0.002), our experimentation shows that the results are not very satisfactory. We plan to propose further testing techniques for programs with very low failure rates.

10. ACKNOWLEDGEMENTS

We are grateful to S.C. Cheung and Chang Xu for their experimental platform as well as their discussions leading to improvements in the paper. We would also like to thank the anonymous reviewers for their invaluable comments.

11. REFERENCES

[1] A. Adi and O. Etzion. Amit: the situation manager. *The VLDB Journal*, 13 (2): 177–203, 2004.

[2] E. W. Axelsen, E. B. Johnsen, and O. Owe. Toward reflective application testing in open environments. In *Proceedings of the Norwegian Informatics Conference* (*NIK 2004*), pages 192–203. Tapir, Trondheim, Norway, 2004.

[3] P. Bellavista, A. Corradi, R. Montanari, and C. Stefanelli. Context-aware middleware for resource management in the wireless Internet. *IEEE TSE*, 29 (12): 1086–1099, 2003.

[4] L. Capra, W. Emmerich, and C. Mascolo. CARISMA: context-aware reflective middleware system for mobile applications. *IEEE TSE*, 29 (10): 929–944, 2003.

[5] A. T. S. Chan and S.-N. Chuang. MobiPADS: a reflective middleware for context-aware mobile computing. *IEEE TSE*, 29 (12): 1072–1085, 2003.

[6] G. Chen and D. Kotz. A survey of context-aware mobile computing research. Technical Report TR2000-381. Dartmouth College, Hanover, New Hampshire, 2000.

[7] A. K. Dey and G. D. Abowd. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-Computer Interaction Journal*, 16 (2-4): 97–166, 2001.

[8] W. Emmerich. Software engineering and middleware: a roadmap. Track on the Future of Software Engineering, in *Proceedings of ICSE 2000*, pages 117–129. ACM Press, New York, 2000.

[9] A. Flores, J. C. Augusto, M. Polo, and M. Varea. Towards context-aware testing for semantic interoperability on PvC environments. In *Proceedings of SMC 2004*, volume 2, pages 1136–1141. IEEE Computer Society Press, Los Alamitos, California, 2004.

[10] P. G. Frankl and O. Iakounenko. Further empirical studies of test effectiveness. In *Proceedings of FSE-6*, pages 153–162. ACM Press, New York, 1998.

[11] P. G. Frankl and S. N. Weiss. An experimental comparison of the effectiveness of branch testing and data flow testing. *IEEE TSE*, 19 (8): 774–787, 1993.

[12] P. G. Frankl and E. J. Weyuker. An applicable family of data flow testing criteria. *IEEE TSE*, 14 (10): 1483–1498, 1988.

[13] M. J. Harrold and M. L. Soffa. Efficient computation of interprocedural definition-use chains. *ACM TOPLAS*, 16 (2): 175–204, 1994.

[14] M. S. Hecht. *Flow Analysis of Computer Programs*. Elsevier, New York, 1977.

[15] M. Hutchins, H. Foster, T. Goradia, and T. Ostrand. Experiments of the effectiveness of dataflow- and controlflow-based test adequacy criteria. In *Proceedings of ICSE 1994*, pages 191–200. IEEE Computer Society Press, Los Alamitos, California, 1994.

[16] Z. Jin and A. J. Offutt. Coupling-based criteria for integration testing. *Software Testing, Verification and Reliability*, 8 (3): 133–154, 1998.

[17] C. Julien and G. C. Roman. Egocentric context-aware programming in ad hoc mobile environments. In *Proceedings of FSE-10*, pages 21–30. ACM Press, New York, 2002.

[18] J. D. Kiper. Structural testing of rule-based expert systems. *ACM TOSEM*, 1 (2): 168–187, 1992.

[19] S. H. Kirani, I. A. Zualkernan, and W.-T. Tsai. Evaluation of expert system testing methods. *CACM*, 37 (11): 71–81, 1994.

[20] G. M. Kapfhammer and M. L. Soffa. A family of test adequacy criteria for database-driven applications. In *Proceedings of ESEC 2003/FSE-11*, pages 98–107, ACM Press, New York, 2003.

[21] H. Lu, W. K. Chan, and T. H. Tse. Testing context-aware middleware-centric programs: a data flow approach and an RFID-based experimentation. Technical Report TR-2006-03, http://www.cs.hku.hk/research/techreps/document/TR-2006-03.pdf, Department of Computer Science, The University of Hong Kong, Pokfulam, Hong Kong, 2006.

[22] A. K. Mok, P. Konana, G. Liu, C.-G. Lee, and H. Woo. Specifying timing constraints and composite events: an application in the design of electronic brokerages. *IEEE TSE*, 20 (12): 841–858, 2004.

[23] A. L. Murphy, G. P. Picco, and G.-C. Roman. LIME: a coordination model and middleware supporting mobility of hosts and agents. *ACM TOSEM*, to appear.

[24] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil. LANDMARC: indoor location sensing using active RFID. *ACM Wireless Networks*, 10 (6): 701–710, 2004.

[25] H. D. Pande, W. A. Landi, and B. G. Ryder. Interprocedural def-use associations for C systems with single level pointers. *IEEE TSE*, 20 (5): 385–403, 1994.

[26] A. Ranganathan and R. H. Campbell. An infrastructure for context-awareness based on first order logic. *Personal and Ubiquitous Computing*, 7 (6): 353–364, 2003.

[27] S. Rapps and E. J. Weyuker. Selecting software test data using data flow information. *IEEE TSE*, SE-11 (4): 367–375, 1985.

[28] M. Román, C. Hess, R. Cerqueira, A. Ranganathan, R. H. Campbell, and K. Nahrstedt. A middleware infrastructure for active spaces. *IEEE Pervasive Computing*, 1 (4): 74–83, 2002.

[29] G.-C. Roman, G. P. Picco, and A. L. Murphy. Software engineering for mobility: a roadmap. Track on the Future of Software Engineering, in *Proceedings of ICSE 2000*, pages 241–258, ACM Press, New York, 2000.

[30] M. Satyanarayanan. Pervasive computing: vision and challenges. *IEEE Personal Communications*, 8 (4): 10–17, 2001.

[31] A. F. Siegel and C. J. Morgan. *Statistics and Data Analysis: an Introduction*. Wiley, New York, 1998.

[32] A. L. Souter and L.L. Pollock. The construction of contextual def-use associations for object-oriented systems. *IEEE TSE*, 29 (11): 1005–1018, 2003.

[33] K.-C. Tai. Theory of fault-based predicate testing for computer programs. *IEEE TSE* 22 (8): 552–562, 1996.

[34] W.-T. Tsai, R. Vishnuvajjala, and D. Zhang. Verification and validation of knowledge-based systems. *IEEE Transactions on Knowledge and Data Engineering*, 11 (1): 202–212, 1999.

[35] T. H. Tse, S. S. Yau, W. K. Chan, H. Lu, and T. Y. Chen. Testing context-sensitive middleware-based software applications. In *Proceedings of COMPSAC 2004*, volume 1, pages 458–465. IEEE Computer Society Press, Los Alamitos, California, 2004.

[36] R. Want, A. Hopper, V. Falcao, and J. Gibbons. The active badge location system. *ACM Transactions on Information Systems*, 10 (1): 91–102, 1992.

[37] C. Xu and S. C. Cheung. Inconsistency detection and resolution for context-aware middleware support. In *Proceedings of ESEC 2005/ FSE-13*, pages 336–345. ACM Press, New York, 2005.

[38] C. Xu, S. C. Cheung, and W. K. Chan. Incremental consistency checking for pervasive context. In *Proceedings of ICSE 2006*, pages 292–301. ACM Press, New York, 2006.

[39] S. S. Yau and F. Karim. An adaptive middleware for context-sensitive communications for real-time applications in ubiquitous computing environments. *Journal of Real-Time Systems*, 26 (1): 29–61, 2004.

[40] H. Zhu, P. A. V. Hall, and J. H. R. May. Software unit test coverage and adequacy. *ACM Computing Surveys*, 29 (4): 366–427, 1997.