Workflow Support for Mobile Data Collection
PHD Research Progress for CAISE 11 Doctoral Consortium

Peter K. Wakholi

University of Bergen, Norway.
peterokhisa@gmail.com,
www.infomedia.uib.no
Supervisor: Assoc Prof: Weiqin Chen

Abstract. Mobile devices have become popular for electronic data collection, management and dissemination in low income countries. Many organizations are considering the use of electronic devices rather than paper-based routines for data collection. In data collection, time dependent and cumulative data, can be regarded as forming a workflow process. Paper-based routines support this workflow process and the transition to a paperless environment poses some workflow challenges that if not addressed could lead to failure to effectively use mobile devices and to meet process-related requirements. The research seeks to explore how challenges related to the use of workflows in mobile, disconnected and distributed environments can be addressed. The research uses an action research approach to propose solutions based on case studies related to Mobile Health (MHealth) and Clinical Trials. This paper presents the current status of research by discussing ideas, methods and frameworks for workflows that provide the flexibility and functionality modeled on paper-based routines.

1 Introduction

In organizations, the daily operation is governed by a set of cooperative business processes, in which interactions with humans and information systems are involved. Workflow systems aim to automate a business processes in whole or part, during which documents, information and tasks are passed from one participant to another for action according to preset procedural rules [1]. Workflow management provides the ability to improve the efficiency of an organization by streamlining and automating coordinated work activities over distributed environments. Organisations using mobile data collection (MDC) aim to achieve this goal by replacing paper-based routines with mobile devices. This research is based on MDC in clinical trials, which provides a good use-case for replacing paper-based routines with mobile devices.

A Clinical Trial can be understood as a workflow process. The workflows are defined based on study protocols, which specify the duration and structure of the study and the standard guidelines which must be followed by all participants [2]. The data collection process is most difficult as it involves field studies which could be in single or multiple sites, often in remote rural environments. One of the core documents in clinical trial for data collection is the Case Report Form (CRF). The CRF is a form where the investigator enters all patients' clinical and non-clinical data related to the trial. There are three types of data: non-time dependent, time dependent and cumulative data [3]. Non-time dependent data is the data collected at a snapshot in time. Such data include subject demographics and medical history. Time dependent data is data collected repeatedly over time through multiple visits. Cumulative data is data collected over time but not linked to a specific visit.

If workflows using mobile-based routines are to replace paper-based routines, some challenges need to be overcome. First, MDC in mobile/wireless computing environment posts new challenges of the such as disconnections, slow connection links, and limited computing power which must be addressed when designing a mobile application. A workflow solution for MDC should take into consideration these characteristics and limitations by employing new computing paradigms. Second, paper-based routines provide the flexibility to work in remote and disconnected environments. Work conducted in the field is often done on paper and filed at the end of a specified time. The workflow system therefore needs to allow for this flexibility by enabling execution of multiple tasks before connection to the server is required. This is a shift from the traditional workflow systems provide for client-server architecture that requires synchronisation before the next task is executed and in some cases web-based-distributed architecture which require constant connection links [4].

This research focuses on the development of mobile systems for EDC by addressing workflow-related problems in their use for clinical trials and mhealth projects. The action research approach is used to develop appropriate methods, models and frameworks for deployment and utilization of Workflow Management System (WFMS) for mobile data collection. The research will be based on the OMEVAC (Open Mobile Electronic Vaccine Trials) project [5] - a new mobile platform for conducting clinical trials, funded by the Norwegian Research Council led by the Centre for International Health, University of Bergen. Case studies will be undertaken by testing and evaluating the prototypes and tools in clinical trials being conducted in rural parts of Africa.

1.1 Research Questions

The main research question is to investigate "how advances in mobile computing, process-aware information systems can combine to enable the use of mobile devices for data collection, management and dissemination in place of paper-based routines used for field activities" In order to achieve new knowledge in this area, the following sub questions will be investigated.

1 How can workflow systems be implemented in the highly constrained and resource-limited mobile environments?
2 How can mobile workflows allow for the flexibility of paper-based routines in field activities?
3 How can distributed workflows be enacted and synchronized in disconnected environments?

1.2 Research Objectives

The research objectives are:-

1 To develop a framework for deployment of generic workflow systems to support mobile data collection in resource constrained environments.
2 To develop methods for execution of multiple tasks before connection to a server is required so as to allow for flexibility mirrored on paper based routines.
3 To test the concepts above by implementing an actual workflow based mobile data collection, management and dissemination system for the OMEVAC project.
4 To demonstrate the value and potential of the mobile data collection in low-income countries as a replacement for paper-based routines.

2 Current Research and Preliminary Ideas

2.1 Workflow Support for Mobile Data Collection

This research (submitted and accepted by CAISE BPMDS 2011) aims to answer the first research question identified. It examines the challenges of deploying (WFMS) in generic data collection tools and proposes a framework for integration. The WFMS is based on YAWL [6] with a generic MDC tool called openXdata [7], [8]. It describes the design and implementation of a workflow adapter that acts as a bridge between the mobile device, data processing applications and workflow engine. Through this implementation, we have been able to provide a distributed architecture that enables the ordering of tasks linked to mobile devices and web-based applications. In addition, we provide an example where this framework has been used in an MHealth project of a vaccination registry that uses mobile devices to collect data on child immunisations.

2.2 Allowing for Offline Behaviour

In order for mobile devices to be used for data collection in disconnected environments, there is need to allow for offline behaviour. This can be achieved by reducing the number of connections required in-between tasks by assigning as many tasks as possible for each connection. Therefore tasks need to be grouped and assigned to a user - assuming that there are no outside data and resource dependencies for execution of the next task.

In WFMS, multiple execution of tasks have been implemented using worklets [9]. Worklets are subnets that are executed like subroutines of a main program and are created at design time. In this research we explore the possibility of dynamically identifying and generating worklets which can then be a assigned as a bundle of tasks to be executed by the mobile user. In order to achieve this, there is need to (1) determine which set of tasks to group; (2) ensure that the tasks assigned are performed as required and (3) synchronise with the server once tasks are completed. In order to determine which set of tasks to group we propose a framework for multiple task assignment that can be used to check the process model to create worklets. To ensure that the worklets are executed on a mobile phone as intended, we propose designing a mobile client to have a light workflow engine that receives a worklet. Tasks in a worklet may be fully or partially completed. We therefore propose a protocol that tracks the status of execution on the mobile client and updates server. This synchronisation protocol would be run every time a connection is established.

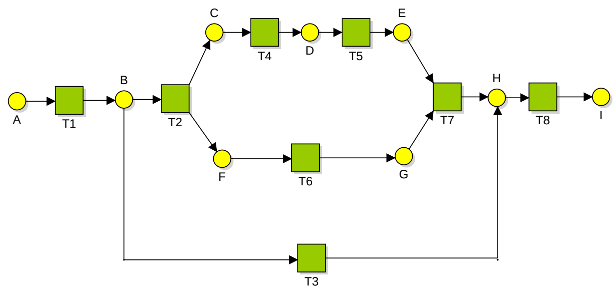

Fig. 1: Running example of a process model

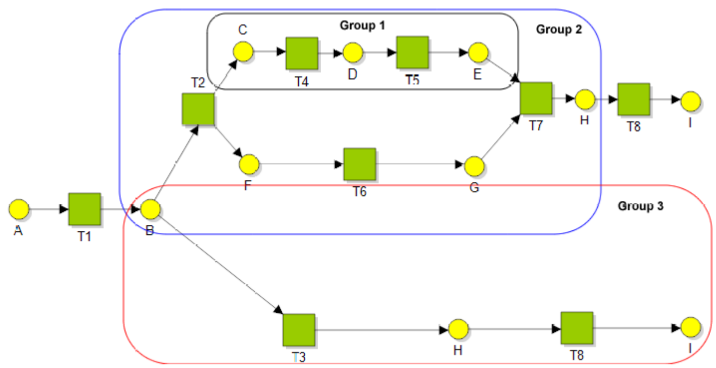

Fig. 2: unfolded process model

3 Theoretical Basis for Proposed Solutions

3.1 Dynamic Generation of Worklets

We propose to use model checking to automatically test a process model and determine whether this model meets a given property. Model checking aims to simulate every execution of the model of the system in order to obtain a labelled reachability graph describing all its behaviors. The graph is the checked against a set of properties to determine if the system performs according to a specified property. We use the model unfolding technique as proposed by [10], [11] to determine the causality, conflicts and concurrency properties. [10] define these properties as; two nodes x and y of a net, are causally retated denoted by $x \leq y$, if there is a path of arrows from x to y. The nodes are in conflict, denoted by $x \neq y$, if there is a place z, different from x and y, from which one can reach x and y, exiting z by different arrows. Figure 2 provides an illustration of this approach.

Figure 1 gives a workflow net showing tasks T1 to T6 that are executed using the basic control flow patterns initially proposed by the WFMC [12] as *Sequence*, *Parallel Split*, *Synchronization*, *Exclusive Choice* and *Simple Merge*. The net presented can be unfolded as shown in 2. The groups 1,2 and 3 present the possible sets of tasks whose grouping would not affect the execution of the patterns above. Based on these groups, the following behavioral observations of the nets can be made:

1 The nodes in group 1 (C,D,E) are in causal relationship and connected by more than one event.There are no arc emerging or entering this group.
2 The nodes in group 2 and 3 (B,C,D,E,G,H) (B,H,I) are a set on a path that has exactly one input condition and one output condition. There are no arcs leaving or entering the group except at node B (the initial node).

From these observations one can conclude that in as long as there exists one input arc and one output arc into a set of nodes from the initial node to the final node, then a worklet is possible. Input or output arcs are permisible on the initial or final node. Therefore given an unfolded model, a worklet can be created from a set of nodes that are not in contradiction with any node outside the set. In addition, for the set of tasks, all the conditions must be causal (i.e. system should transition to another state in the group) and all events must be strongly connected (i.e. There exists a path from first event to the last event).

Proposition 1. *Given an unfolding of a workflow net $A = \langle S; T; \alpha; \beta; is \rangle$, such that $S = \{s_0, s_1, ..., s_m\}$ is a set of states and $T = \{t_0, t_1, ..., t_n\}$ is a set of transitions, a worklet $W = \langle S'; T'; \alpha; \beta; is \rangle$ where $T' = \{t_i, t_{i+1}, ...t_{i+m}\} \subset T$ and $S' = \{s_i, s_{i+1}, ...s_{i+n}\} \subset S$ is possible if:*

1. $\forall S' \exists \{s_i \leq s_{i+1}\}$ *where* $0 \leq i \leq m$ *(Causality)*
2. $\forall S' \exists s_i$ *such that* $s_i \neg \neq \{S' \notin S\}$ *(no conflict)*

3.2 Event-log based Synchronization

Synchronisation takes place once a connection between a server and mobile client is established. Any synchronisation solution should aim to ensure appropriate utilisation mobile phone processing and memory. We therefore propose keeping a log of all the events on a mobile device and synchronising with the server. For synchronisation, reachability analysis can be used to compare the workflow net and event log. For a transition system like a workflow net, we provide a definition of a state called *reached marking* as a basis for tracking current work status.

Definition 1 ***Reached Marking:*** *Let p be a state in P and $\alpha = (a_1, a_2.......a_m) \in A$ be a set of executed tasks for the process definition of P. State p_0 is the initial state of the process and p_n is the final state. α are reached markings if and only if there exists states $s_0, s_1, ..., s_m$ such that $p_0 = s_0$ and $s_i \xrightarrow{a_i+1} s_{i+1}$ for all for all $0 \leq i \leq m$. The next execution task satisfies the relation $s_m \xrightarrow{a_m+1} s_{m+1}$ and $p_n \neq s_{m+1}$*

Figure 3 illustrates an example of a simple workflow for patient visiting a health clinic in a clinical trial. The reachability graph shows all the possible states of the system and transitions that transform these states. Note that the states associated with the

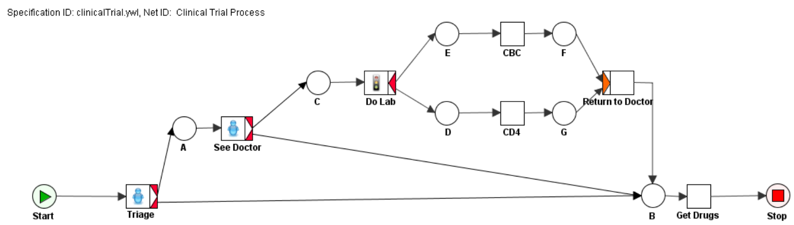

Fig. 3: Synchronisation working example

PatientID	Task	End-time
1	Triage/ Do get Drugs	12:00
1	Get Drugs	12:10
2	Triage/Do see Dr	12:11
3	Triage/Do see	12:30
1	Get Drugs	12:10
2	See Doctor	13:00
2	Do Lab	13:10
2	CD4	13:15

Table 1: event log

hidden tasks are painted black because in actual sense they do not exist. Assume that part of this process exists on the server and a client implementing it as a workflet. The client keeps an event log indicating the tasks executed and the end-time of execution. Suppose the event log is as illustrated in the table below:

We now propose an algorithm to determine the reached markings algorithm. Throughout the algorithm a set of vertices, which corresponds to state of work for each case, denoted by P is maintained. Initially, each node is unmarked. A node P_i is marked means that a task s_i has been executed and node P_i in P has been reached. Initially, only P_0 is marked. Using a sequential set of event logs for process A; for each node, check for the log to determine the next state. While a task exists and corresponds to an edge from the current state, then mark the next node and remove task from log. The last marked node corresponds to the current state of work. If the state is not the end node, the outgoing edge from that state corresponds to next task to be undertaken. If the node is an end node P_n, then the case is complete. The following algorithm provides the pseudo code for determining the reached markings:

int $count = 0$
set $\alpha = (a_1, a_2.......a_m)$
node $initialNode = p_n$
node $finalNode = p_n$
reachability (graph P, node p_i, edge s_i)
mark p_0
while $\alpha \neq \emptyset$ **do**

```
    for all $a_i$ do
      for all $p_i$ do
        if match $(s_i, a_i)$ then
          mark $p_i$
        end if
      end for
    end for
    $reachedNode = p_i$
  end while
  if $reachedNode \neq p_n$ then
    Return $p_{i+1}$
  else
    Return $p_n$
  end if
```

4 Conclusion

We have presented in this paper ongoing research about workflow support for mobile data collection. The use of worklets to enable execution of multiple tasks on a mobile device reduces the need for many connections thereby allowing for more flexibility in a disconnected environment. Moreover, it would be neccessary to have these worklets dynamically generated in order impose less restrictions on the system. We proposed a model checking technique based on model unfolding and genetic algorithms to generate these worklets. In order to enable synchronisation between the workflow server and worklet client, we propose event-log based synchronisation. The use of this approach greatly reduces the memory usage and application footprint - a key requirement for the limited resource mobile environment. We thus provide clear progress towards achieving the objectives of the research as highlighted by the practical ideas and theory presented. The mobile data collection tool will be enhanced to allow for offline behaviour using these ideas. It is anticipated that all methods and frameworks proposed will be empirically tested in MHealth projects to validate the research.

References

1. M. Weske. *Business process management: concepts, languages, architectures*. Springer-Verlag New York Inc, 2007.
2. US Food. Drug Administration. Guidance for industry, E6 good clinical practice: consolidated guidance. *Federal Register*, 10:691–709, 1997.
3. K.K. Moon. Techniques for Designing Case Report Forms in Clinical Trials. *ScianNews Volume, DOI*, 2006.
4. R.S.H. Istepanian and C.S. Pattichis. *M-health: Emerging mobile health systems*. Springer-Verlag New York Inc, 2006.
5. J. Klungsoyr, T. Tylleskar, B. Macleod, P. Bagyenda, W. Chen, and P. Wakholi. OMEVAC–Open Mobile Electronic Vaccine Trials, an interdisciplinary project to improve quality of vaccine trials in low-resource settings. *M4D 2008, General Tracks*.

6. W.M.P. Van Der Aalst and A.H.M. Ter Hofstede. YAWL: yet another workflow language. *Information Systems*, 30(4):245–275, 2005.
7. Y. Anokwa, C. Hartung, A. Lerer, B. DeRenzi, and G. Borriello. A new generation of open source data collection tools. In *Information and Communication Technologies and Development (ICTD), 2009 International Conference on*, page 493. IEEE, 2010.
8. J. Klungsoyr, P. Wakholi, B. Macleod, A. Escudero-Pascual, and N. Lesh. OpenROSA, JavaROSA, GloballyMobile–Collaborations around Open Standards for Mobile Applications. *M4D 2008, General Tracks*.
9. M. Adams, A. ter Hofstede, D. Edmond, and W. van der Aalst. Worklets: A service-oriented implementation of dynamic flexibility in workflows. *On the Move to Meaningful Internet Systems 2006: CoopIS, DOA, GADA, and ODBASE*, pages 291–308, 2006.
10. J. Esparza and K. Heljanko. *Unfoldings: a partial-order approach to model checking*. Springer-Verlag New York Inc, 2008.
11. J. Esparza and K. Heljanko. Implementing LTL model checking with net unfoldings. *Model Checking Software*, pages 37–56, 2001.
12. D. Hollingsworth. Workflow Management Coalition-Terminology & Glossary,[online], 1999, verf
"ugbar im World Wide Web: http://www. wfmc. org/standards/docs. *TC-1011_term_glossary_v3. pdf*.