XRPC: efficient distributed query processing on heterogeneous XQuery engines
Zhang, Y.

Citation for published version (APA):
Zhang, Y. (2010). XRPC: efficient distributed query processing on heterogeneous XQuery engines

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
# Contents

1 Introduction

1.1 Motivation

1.1.1 Interoperability

1.1.2 Extending XQuery with Query Shipping

1.1.3 Efficiency

1.1.4 Stateless versus Stateful

1.2 Research Objective

1.3 Thesis Outline

2 Related Work

2.1 P2P Data Management Systems

2.1.1 Extending XQuery with Query-Shipping

2.1.2 Distributed XML Querying in Structured P2P Systems

2.1.3 Distributed XML Querying in Unstructured P2P Systems

2.1.4 PDMSs of Relational Data

2.2 Related Query Processing Techniques

2.2.1 XML Document Filtering

2.2.2 Query Decomposition

2.3 Conclusion

3 The XRPC Language Extension

3.1 Design Considerations

3.2 XRPC Syntax

3.3 SOAP XRPC Message Format

3.3.1 XRPC Request Messages

3.3.2 XRPC Response Messages

3.3.3 XRPC Error Message

3.4 XRPC Formal Semantics

3.4.1 Read-Only XRPC Semantics

3.4.2 XRPC Update Semantics

3.5 Loop-lifted Implementation of XRPC

3.5.1 Relational XQuery and Loop-Lifting

3.5.2 Bulk RPC

3.5.3 Performance Evaluation

3.6 Conclusion
## 4 Distributed XQuery With XPRC 45

### 4.1 Introduction 45

### 4.2 Cross-System Distributed XQuery 47

### 4.3 Distributed XQuery Optimisation 49

#### 4.4 Deterministic Distributed Updates 51

##### 4.4.1 Order-Correct Update Tags 52

### 4.4.1 Heterogeneous Distributed 2PC 56

### 4.5 Distributed XRPC Transactions 55

##### 4.5.1 Heterogeneous Distributed 2PC 56

### 4.6 MonetDB/XQuery 58

#### 4.6.1 Simple Scenarios 58

#### 4.6.2 Loose DHT Coupling 58

#### 4.6.3 Tight DHT Coupling 60

### 4.7 Conclusion 60

## 5 XQuery Decomposition 63

### 5.1 Motivation 63

### 5.2 Semantic Differences with Pass-By-Value 65

### 5.3 XQuery Core Rewrite Framework 68

#### 5.3.1 XCore Dependency Graph 69

#### 5.3.2 XRPCExpr Insertion 71

### 5.4 Conservative Decomposition 71

#### 5.4.1 By-Value Insertion Conditions 71

#### 5.4.2 Interesting Decomposition Points 72

#### 5.4.3 Normalisation 74

#### 5.4.4 Distributed Code Motion 74

### 5.5 By-Fragment Decomposition 75

### 5.6 By-Projection Decomposition 77

#### 5.6.1 Extending Projected XML 79

#### 5.6.2 Runtime XML Projection 81

### 5.7 Decomposition of XQUF Queries 84

#### 5.7.1 Distributing Normal XQUF Queries 84

#### 5.7.2 Updating XCore Queries on Remote Documents 86

### 5.8 Evaluation in MonetDB/XQuery 89

#### 5.8.1 Read-Only Queries 89

#### 5.8.2 XQUF Queries 91

### 5.9 Conclusion 94

## 6 Correctness Proof of XQuery Decomposition 95

### 6.1 Preliminaries 95

#### 6.1.1 Equality Relationships of Sequences 96

#### 6.1.2 Equality Relationships of Sequences with Projection 96

#### 6.1.3 Equality Relationship of Read-Only Queries 98

#### 6.1.4 Equality Relationship of Updating Queries 100

#### 6.1.5 Sequence Properties 102

#### 6.1.6 XPath Steps and `distinct-doc-order` 103

### 6.2 Static Properties Analysis 104