



Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

Binary Serialization of XML for Small Devices

Jaakko Kangasharju
jkangash@cs.helsinki.fi

University of Helsinki

Ph.D. Student Workshop on Spontaneous Networking,
Rutgers University, May 2006



Outline

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- 1 Introduction
- 2 General-Purpose Techniques
- 3 Schema-Based Techniques
- 4 Complete Xebu Example
- 5 Measurements
- 6 Conclusions



Outline

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- 1 Introduction
- 2 General-Purpose Techniques
- 3 Schema-Based Techniques
- 4 Complete Xebu Example
- 5 Measurements
- 6 Conclusions



Messaging and Message Formats

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- Messaging: sending and receiving pieces of structured typed data
- Message format specifies how structure is represented
- Small mobile device requirements: small footprint, efficient processing, compact
- Possible formats: ASN.1, CORBA CDR, Java serialization, custom formats
- On fixed networks XML a recent text-based flexible alternative
- Large variety of XML-based technologies make it attractive



XML Example

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

XML:

```
<?xml version="1.0" encoding="UTF-8"?>
<p:person nation="DE" xmlns:p="http://example.org/people">
  <p:name>
    <p:first>Richard</p:first>
    <p:last>Wagner</p:last>
  </p:name>
  <p:occupation>Composer</p:occupation>
  <p:born>1813-05-22</p:born>
  <p:died>1883-02-13</p:died>
</p:person>
```



XML Example

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

XML:

```
<?xml version="1.0" encoding="UTF-8"?>
<p:person nation="DE" xmlns:p="http://example.org/people">
  <p:name>
    <p:first>Richard</p:first>
    <p:last>Wagner</p:last>
  </p:name>
  <p:occupation>Composer</p:occupation>
  <p:born>1813-05-22</p:born>
  <p:died>1883-02-13</p:died>
</p:person>
```

Unix passwd style:

```
Richard::Wagner:DE:Composer:1813-05-22:1883-02-13
```



Binary XML Formats

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- Compression not always feasible
- **Binary XML** format to avoid string processing
- Compatibility with XML processing APIs
- Need direct readability and writability without going through XML
- Several existing formats: WAP Binary XML, Fast Infoset, XBIS, Xebu, XSBC, BiM, . . .



Outline

Binary XML

Jaakko
Kangasharju

Introduction

**General-
Purpose
Techniques**

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- 1 Introduction
- 2 General-Purpose Techniques**
- 3 Schema-Based Techniques
- 4 Complete Xebu Example
- 5 Measurements
- 6 Conclusions



Requirements

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- To replace XML, must be able to represent any XML
- Must achieve higher compactness and processing efficiency
- Typically based on some XML **data model** (Infoset, XPath, SAX, ...)
- Fast Infoset based on Infoset, Xebu based on XAS



Tokenization

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- Element and attribute names repeat a lot in XML
- **Tokenize** names (and possibly content) by replacing them with small binary values
- Do not repeat at end tags
- Dynamic tokenization general-purpose
- Tokenizing formats: Fast Infoset, XBIS, Xebu



Outline

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- 1 Introduction
- 2 General-Purpose Techniques
- 3 Schema-Based Techniques**
- 4 Complete Xebu Example
- 5 Measurements
- 6 Conclusions



Schema Languages

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

Schema languages describe structure and content of XML documents (DTD, XML Schema, RELAX NG)

```
namespace p = "http://example.org/people"
start = element p:person {
  attribute nation { xsd:token { pattern = "\w\w" } }?,
  element p:name {
    element p:first { token },
    element p:middle { token }?,
    element p:last { token }
  },
  element p:occupation { token }?,
  element p:born { xsd:date },
  element p:died { xsd:date }
}
```



Simple Techniques

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- Information given by schema usable to achieve improved compactness
- Use of schema may tie serialized form to schema
- For tokenizing formats, pre-tokenize strings appearing in schema
- Use data type information to select more efficient encodings
- Formats using these techniques: Fast Infoset, XSBC, Xebu



The XAS Data Model

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- Advanced schema-based technique of Xebu based on XAS data model
- XML document a sequence of **events**, each containing strings

XML:

```
<p:person nation="DE" xmlns:p="http://example.org/people">  
  ...  
</p:person>
```

Event sequence:

```
PM(p=http://example.org/people) SE(p:person)  
A(nation=DE) C(...) EE(p:person)
```



Event Omission

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- Construct two **event omission automata**, for serialization and parsing
- Transitions on serialization side can omit events
- Transitions on parsing side insert omitted events back
- Letting unrecognized events through allows some deviations from schema
- Working on event sequence level achieves independence from serialization format



Outline

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- 1 Introduction
- 2 General-Purpose Techniques
- 3 Schema-Based Techniques
- 4 Complete Xebu Example
- 5 Measurements
- 6 Conclusions



Example Schema and Document

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

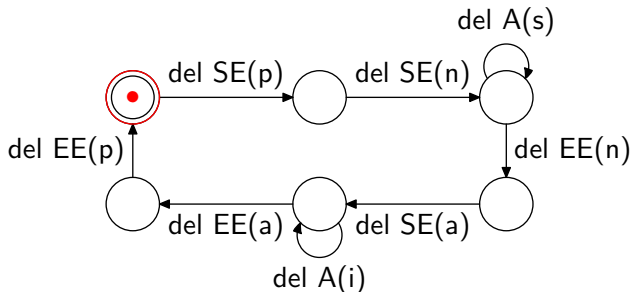
Conclusions

Schema	Document
<pre>start = element person { element name { xsd:string }, element age { xsd:int } }</pre>	<pre><person> <name>Alice</name> <work>Research</work> <age>30</age> </person></pre>
Event sequence	
<pre>SE(person) SE(name) TC(string,Alice) EE(name) SE(work) C(Research) EE(work) SE(age) TC(int,30) EE(age) EE(person)</pre>	
Initial token mapping	
<pre>NS""=0, N"{}person"=0, N"{}name"=1, N"{}age"=2</pre>	



Example Serialization

▲ SE(person) SE(name) A(type=string)
TC(string, Alice) EE(name) SE(work) C(Research)
EE(work) SE(age) A(type=int) TC(int, 30)
EE(age) EE(person)



Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions



Example Serialization

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

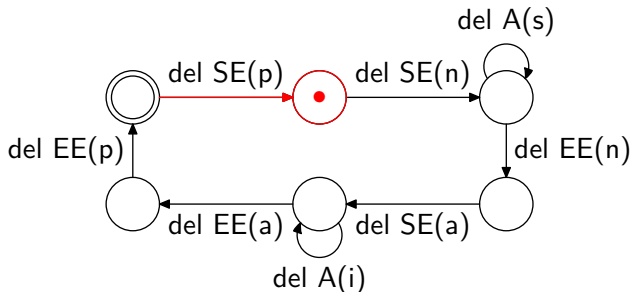
Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

SE(person) SE(name) A(type=string)
TC(string, Alice) EE(name) SE(work) C(Research)
EE(work) SE(age) A(type=int) TC(int, 30)
EE(age) EE(person)





Example Serialization

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

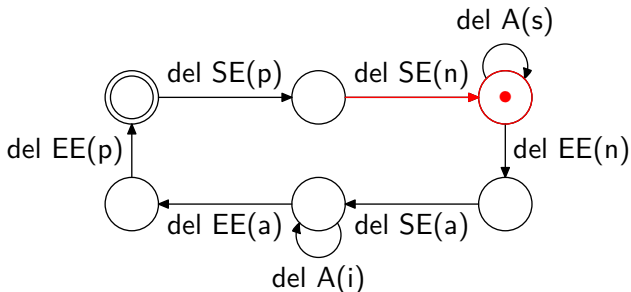
Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

SE(person) SE(name) A(type=string)
TC(string, Alice) EE(name) SE(work) C(Research)
EE(work) SE(age) A(type=int) TC(int, 30)
EE(age) EE(person)





Example Serialization

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

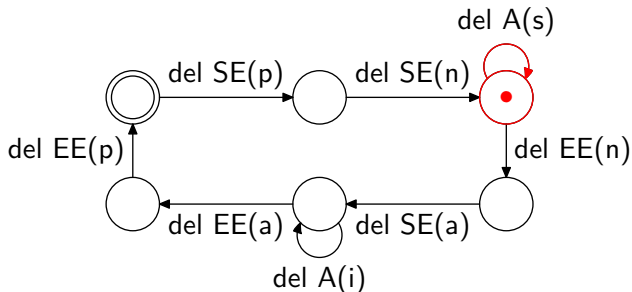
Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

SE(person) SE(name) A(type=string)
TC(string,Alice) EE(name) SE(work) C(Research)
▲ EE(work) SE(age) A(type=int) TC(int,30)
EE(age) EE(person)





Example Serialization

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

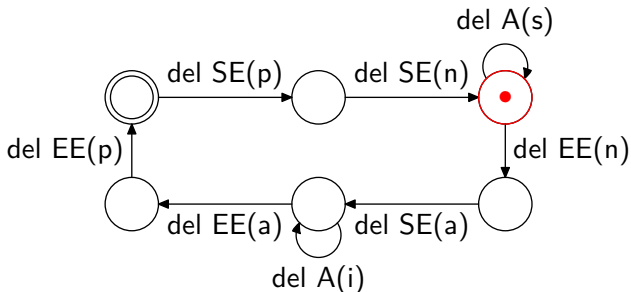
Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

SE(person) SE(name) A(type=string)
TC(string, Alice) EE(name) SE(work) C(Research)
EE(work) SE(age) A(type=int) TC(int, 30)
EE(age) EE(person)



<0C><05>Alice



Example Serialization

Binary XML

Jaakko Kangasharju

Introduction

General-Purpose Techniques

Schema-Based Techniques

Complete Xebu Example

Measurements

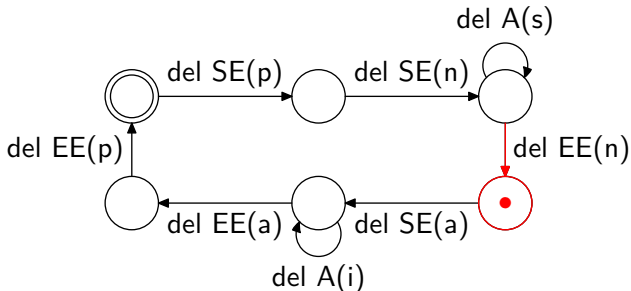
Conclusions

SE(person) SE(name) A(type=string)

TC(string,Alice) EE(name) SE(work) C(Research)

EE(work) SE(age) A(type=int) TC(int,30)

EE(age) EE(person)



<0C><05>Alice



Example Serialization

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

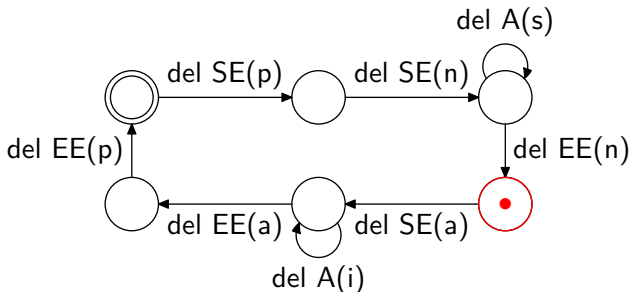
Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

SE(person) SE(name) A(type=string)
TC(string,Alice) EE(name) SE(work) C(Research)
EE(work) SE(age) A(type=int) TC(int,30)
EE(age) EE(person)



<0C><05>Alice

<43><00><03><04>work



Example Serialization

Binary XML

Jaakko Kangasharju

Introduction

General-Purpose Techniques

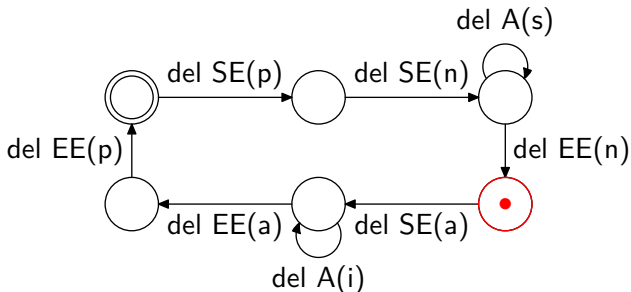
Schema-Based Techniques

Complete Xebu Example

Measurements

Conclusions

SE(person) SE(name) A(type=string)
TC(string,Alice) EE(name) SE(work) C(Research)
EE(work) SE(age) A(type=int) TC(int,30)
▲ EE(age) EE(person)



<0C><05>Alice

<43><00><03><04>work

<07><08>Research



Example Serialization

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

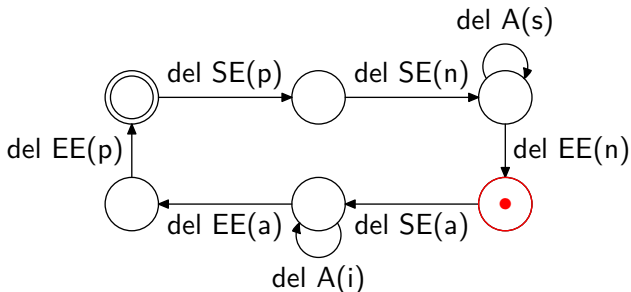
Conclusions

SE(person) SE(name) A(type=string)

TC(string,Alice) EE(name) SE(work) C(Research)

EE(work) SE(age) A(type=int) TC(int,30)

EE(age) EE(person)



<0C><05>Alice

<43><00><03><04>work

<07><08>Research

<04>



Example Serialization

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

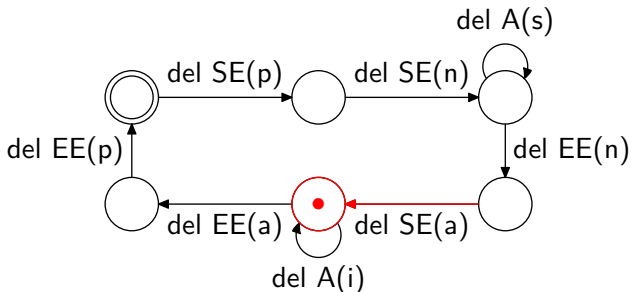
Conclusions

SE(person) SE(name) A(type=string)

TC(string,Alice) EE(name) SE(work) C(Research)

EE(work) SE(age) A(type=int) TC(int,30)

EE(age) EE(person)



<0C><05>Alice

<43><00><03><04>work

<07><08>Research

<04>



Example Serialization

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

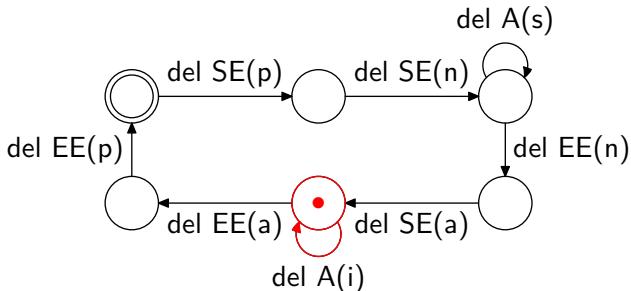
Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

SE(person) SE(name) A(type=string)
TC(string,Alice) EE(name) SE(work) C(Research)
EE(work) SE(age) A(type=int) ▲ TC(int,30)
EE(age) EE(person)



<0C><05>Alice

<43><00><03><04>work

<07><08>Research

<04>



Example Serialization

Binary XML

Jaakko Kangasharju

Introduction

General-Purpose Techniques

Schema-Based Techniques

Complete Xebu Example

Measurements

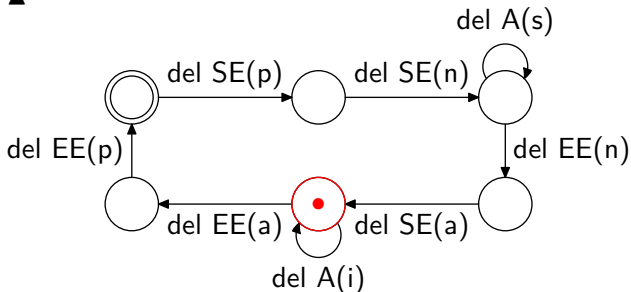
Conclusions

SE(person) SE(name) A(type=string)

TC(string,Alice) EE(name) SE(work) C(Research)

EE(work) SE(age) A(type=int) TC(int,30)

▲ EE(age) EE(person)



<0C><05>Alice

<43><00><03><04>work

<07><08>Research

<04>

<0C><1E>



Example Serialization

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

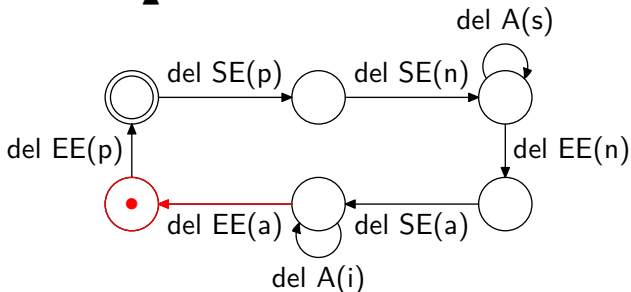
Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

SE(person) SE(name) A(type=string)
TC(string,Alice) EE(name) SE(work) C(Research)
EE(work) SE(age) A(type=int) TC(int,30)
EE(age) ▲ EE(person)



<0C><05>Alice

<43><00><03><04>work

<07><08>Research

<04>

<0C><1E>

Example Serialization

Binary XML

Jaakko Kangasharju

Introduction

General-Purpose Techniques

Schema-Based Techniques

Complete Xebu Example

Measurements

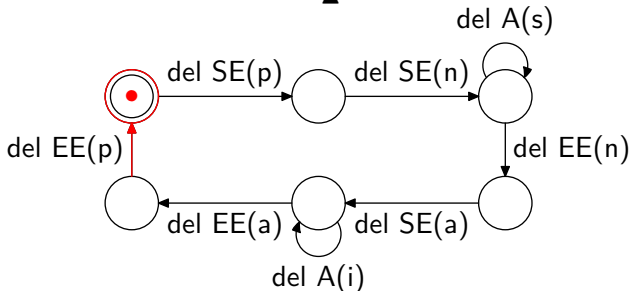
Conclusions

SE(person) SE(name) A(type=string)

TC(string,Alice) EE(name) SE(work) C(Research)

EE(work) SE(age) A(type=int) TC(int,30)

EE(age) EE(person) ▲



<0C><05>Alice

<43><00><03><04>work

<07><08>Research

<04>

<0C><1E>



Example Parsing

Binary XML

Jaakko
Kangasharju

Introduction

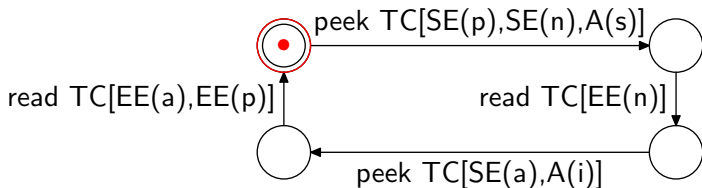
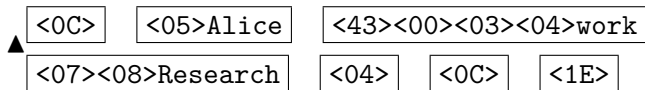
General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions





Example Parsing

Binary XML

Jaakko
Kangasharju

Introduction

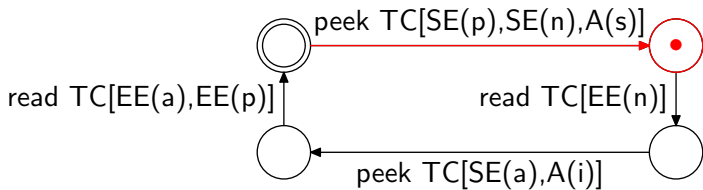
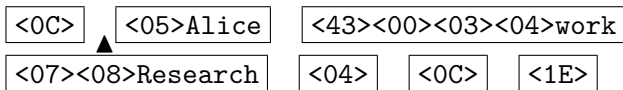
General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions



SE(person) SE(name) A(type=string)



Example Parsing

Binary XML

Jaakko Kangasharju

Introduction

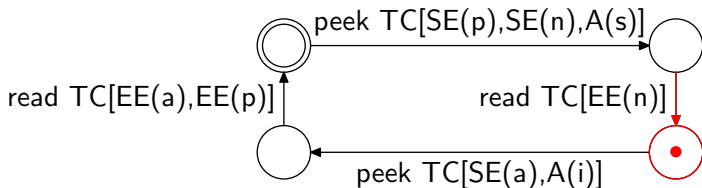
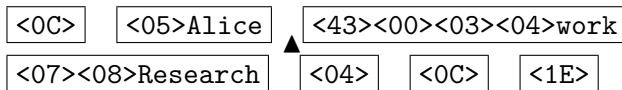
General-Purpose Techniques

Schema-Based Techniques

Complete Xebu Example

Measurements

Conclusions



SE(person) SE(name) A(type=string) TC(string, Alice)
EE(name)



Example Parsing

Binary XML

Jaakko Kangasharju

Introduction

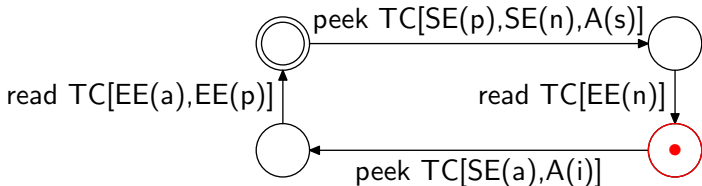
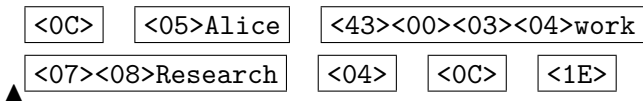
General-Purpose Techniques

Schema-Based Techniques

Complete Xebu Example

Measurements

Conclusions



SE(person) SE(name) A(type=string) TC(string, Alice)
EE(name) SE(work)



Example Parsing

Binary XML

Jaakko
Kangasharju

Introduction

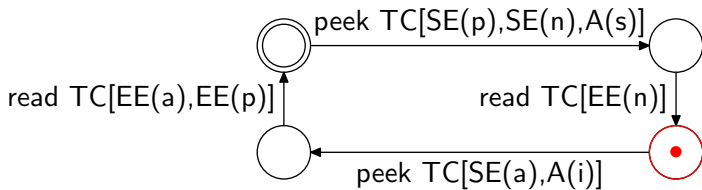
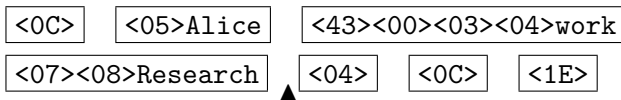
General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions



$SE(\text{person})$ $SE(\text{name})$ $A(\text{type}=\text{string})$ $TC(\text{string}, \text{Alice})$
 $EE(\text{name})$ $SE(\text{work})$ $C(\text{Research})$



Example Parsing

Binary XML

Jaakko Kangasharju

Introduction

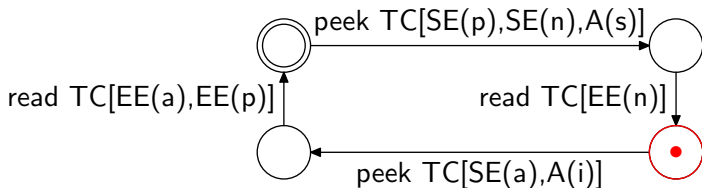
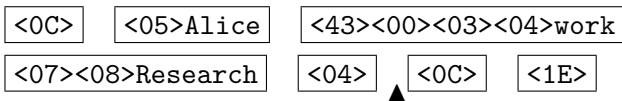
General-Purpose Techniques

Schema-Based Techniques

Complete Xebu Example

Measurements

Conclusions



`SE(person) SE(name) A(type=string) TC(string, Alice)`
`EE(name) SE(work) C(Research) EE(work)`



Example Parsing

Binary XML

Jaakko
Kangasharju

Introduction

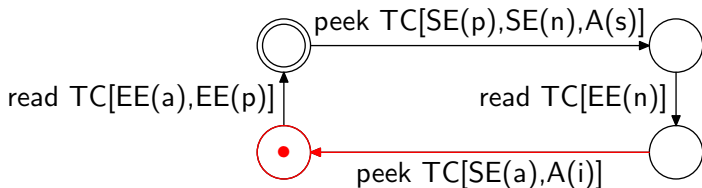
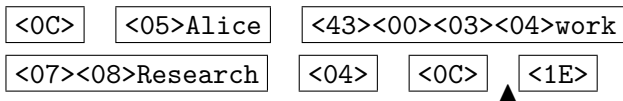
General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions



SE(person) SE(name) A(type=string) TC(string, Alice)
EE(name) SE(work) C(Research) EE(work) SE(age)
A(type=int)



Example Parsing

Binary XML

Jaakko
Kangasharju

Introduction

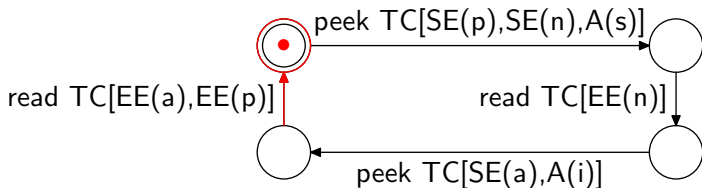
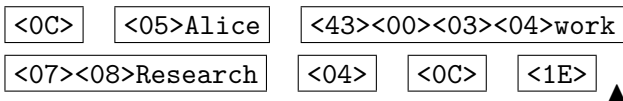
General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions



SE(person) SE(name) A(type=string) TC(string, Alice)
EE(name) SE(work) C(Research) EE(work) SE(age)
A(type=int) TC(int, 30) EE(age) EE(person)



Outline

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- 1 Introduction
- 2 General-Purpose Techniques
- 3 Schema-Based Techniques
- 4 Complete Xebu Example
- 5 Measurements**
- 6 Conclusions



Measurement Setup

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- Desktop machine: 3 GHz Pentium 4, Debian GNU/Linux, Sun Java Development Kit 5.0, heap size 256 MiB
- 698 SOAP messages from an event notification application
- Implementations: Xerces, kXML, gzipped Xerces, Fast Infoset, XBIS, Xebu
- Only tokenization for binary XML formats
- Not directly translatable to small devices



Size Results

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

Format	Size (B)	Size (%)	Foot (KB)
Xerces	3033	100.0	521
kXML	3041	100.3	49
gzip	675	22.3	–
FI	1689	55.7	197
XBIS	1635	53.9	58
Xebu	1314	43.3	58



Processing Results

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

Format	Serialization		Parsing	
	Time (ms)	Mem (KB)	Time (ms)	Mem (KB)
Xerces	0.21	32.3	0.41	104.6
kXML	0.86	137.1	0.69	87.4
gzip	0.51	40.7	0.57	104.9
FI	0.10	16.2	0.15	30.8
XBIS	0.12	19.9	0.16	25.9
Xebu	0.25	26.2	0.32	61.6



Outline

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- 1 Introduction
- 2 General-Purpose Techniques
- 3 Schema-Based Techniques
- 4 Complete Xebu Example
- 5 Measurements
- 6 Conclusions



Conclusions

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions



Conclusions

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- XML performance improvements mostly irrelevant



Conclusions

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- XML performance improvements mostly irrelevant
- **Binary XML is viable**



Conclusions

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- XML performance improvements mostly irrelevant
- Binary XML is viable
- **Format work basically done**



Conclusions

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

- XML performance improvements mostly irrelevant
- Binary XML is viable
- Format work basically done
- **Integration work only starting**



Thank You

Binary XML

Jaakko
Kangasharju

Introduction

General-
Purpose
Techniques

Schema-Based
Techniques

Complete
Xebu Example

Measurements

Conclusions

Questions?