Materialized Views in a Distributed Event Stream Processing Environment

Products

- **Metadata repository and service**
  The metadata repository is stored in an Oracle database with a service-based front-end implemented using the Windows Communication Framework. The implementation includes a collection of C# classes that allow access to the metadata of the various data sources and queries. Details regarding the metadata repository and its services are available in the research paper presented at the SEDE2010 conference in June 2010.

- **CORAL8 CCL Antlr 3.0 Parser**
  Using an initial Antlr grammar that was supplied by the Coral8 technical support, the research group revised the grammar to obtain detailed metadata on the query expressions within the continuous queries of CCL. This information is required for the later identification of common subexpressions.

- **Common Subexpression Algorithm**
  This is a C# implementation of the common subexpression algorithm that takes a mixed multigraph model representation of the various query expressions such as SQL, XQuery and LINQ as the input and generates a list of identical and subsumed selection and join statements. This algorithm is used in the DEPA prototype to identify the common subexpressions as the potential candidates for hybrid materialized views. The algorithm has been evaluated over simple and complex multigraphs over both the Criminal Justice and TPC-H enterprises.

- **Criminal Justice Enterprise**
  The Criminal Justice enterprise is a database enterprise that is based on more complex and realistic scenarios from GJXDM. The details about the enterprise are shown in the UML diagram. The Criminal Justice enterprise consists of SQL scripts to generate relational tables, XML schemas for the relevant XML data sources, and a data generation program. The data generation program is written using C# and LINQ. The program generates comma-separated files, which can be imported into the relational databases and XML files for the XML data sources.

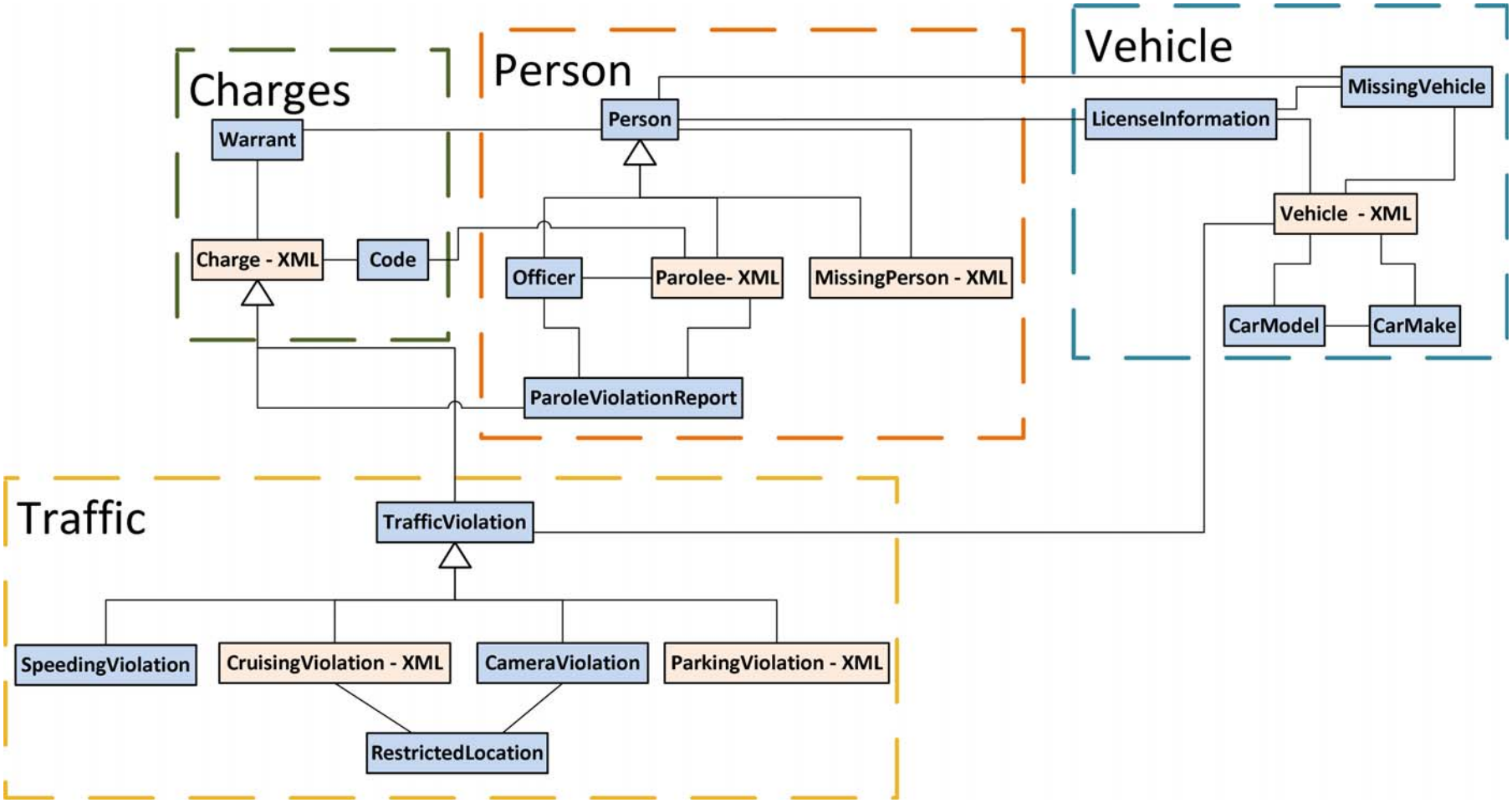

Figure 1. Criminal Justice UML Diagram

- **Simple proof-of-concept Criminal Justice Scenario**
  This is a simple implementation of a small Criminal Justice Scenario that illustrates the various features of the DEPA framework. The example scenario includes data sources stored in Oracle and SQL Server relational databases and XML documents, events, streams, and a LINQ view over a relational and XML source. Research paper illustrating the features of the framework and the details on the proof-of-concept was presented at the VLDB2010 Ph.D. workshop.

- **Sample hybrid scenarios over Criminal Justice enterprise**
  In order to test the working of the different algorithms within the DEPA framework, the research group has focused on designing hybrid

scenarios using the Criminal Justice enterprise. These hybrid scenarios consists of queries defined in SQL, LINQ and XQuery over various heterogeneous data sources. These queries are used in the evaluation of the common subexpression detection algorithm and materialized view definition algorithm in a simulated DEPA framework. The sample scenarios also include the deltas which are randomly generated to simulate the changes in the data sources. The delta generation program is written in C# and uses LINQ to communicate with the database. These deltas are streamed within the DEPA environment to the materialized views and the incremental view maintenance algorithm uses these deltas to update the views.

- **DBGEN for TPC-H scale factor < 1GB**
  The TPC provides the DBGEN software for generating benchmark data starting at 1GB and higher scale factors. This program has been modified to generate benchmark data that is less than 1GB. This item also includes queries to validate the resulting instance according to the benchmark specification.

- **LINQ to SQL Queries for TPC-H**
  The 22 queries of the original, relational TPC-H benchmark written in LINQ to SQL.

- **Data and Queries for a Hybrid TPCH Benchmark over Heterogeneous Data Sources**
  A proposed benchmark that supports queries over heterogeneous data sources (relational and XML), which is based on the well-known TPC-H benchmark.

- **Incremental View Maintenance Prototype**
  The incremental view maintenance prototype illustrates the use of extended magic sets to propagate the relevant deltas to update the hybrid materialized views. This prototype also uses magic sets to materialize the views for the first time. The C# project includes scenarios over both the Criminal Justice and TPC-H enterprises.