New Data Types and Operations to Support Geo-streams

Yan Huang    Chengyang Zhang

Department of Computer Science and Engineering
University of North Texas

GIScience 2008
Outline

1 Motivation
   - Introduction to Geo-stream
   - Motivation Example

2 Previous Work
   - Previous Work

3 Our Contributions
   - Streaming Data Types
   - Geo-stream Operations
   - Language Embeddings
What is Geo-stream?

**Definition**

A geo-stream refers to the **moving** and **changing** geometry of an object, which is continuously fed into a **geo-spatial data stream management system** in **real time**.

**Example**

- **Point** geo-stream: real time location of a moving object
- **Region** geo-stream: evolving spatial extent of a hurricane
What is Geo-stream?

**Definition**
A geo-stream refers to the moving and changing geometry of an object, which is continuously fed into a geo-spatial data stream management system in real time.

**Example**
- **Point** geo-stream: real time location of a moving object
- **Region** geo-stream: evolving spatial extent of a hurricane
Many sensors are geo-referenced.

Figure: Example - MSR SenseWeb Project
Many sensors are geo-referenced.

Figure: Example - Texas Environmental Observation
Many sensors are geo-referenced.

Figure: Example - 3D Google Weather
Introduction to Geo-stream

- Real-time monitoring and alerting need support of querying streaming spatio-temporal extents.

Figure: Example - San Diego Fire Alert Map
Real-time monitoring and alerting need support of querying streaming spatio-temporal extents.

Figure: Example - Yahoo Map
Outline

1. Motivation
   - Introduction to Geo-stream
   - Motivation Example

2. Previous Work
   - Previous Work

3. Our Contributions
   - Streaming Data Types
   - Geo-stream Operations
   - Language Embeddings
Example

- Q1: Notify me when my house is within 50 miles of the mandatory evacuation area of a forest fire.
Motivation Example - Hazard Weather Notification

Example

Q2: Continuously list the addresses of all the houses traversed by flood in the past 2 days in Denton county.
Motivation Example - Hazard Weather Notification

Example

Q3: Continuously list road segments that have been completely under flood-water for the past 24 hours.
Motivation

1. Introduction to Geo-stream
2. Motivation Example

Previous Work

2. Previous Work

Our Contributions

3. Streaming Data Types
4. Geo-stream Operations
5. Language Embeddings

Outline
Spatio-temporal databases (data-type-based approach [5, 7, 12, 8, 6])

- Elegant model to support spatio-temporal domain applications
- Data types, operations, predicates, and language embeddings have been precisely defined
- However, real time geo-streams are not supported

Moving Object Databases [9, 11, 10]

- Large body of work on indexing and query optimization on moving objects
- Limited support of spatio-temporal objects (mostly points)
- Only locations are dynamic, not monitored phenomena
Previous Work

Spatio-temporal databases (data-type-based approach [5, 7, 12, 8, 6])
- Elegant model to support spatio-temporal domain applications
- Data types, operations, predicates, and language embeddings have been precisely defined
- However, real time geo-streams are not supported

Moving Object Databases [9, 11, 10]
- Large body of work on indexing and query optimization on moving objects
- Limited support of spatio-temporal objects (mostly points)
- Only locations are dynamic, not monitored phenomena
## Previous Work

### Data Stream Management System (DSMS)\[4, 13, 3, 2, 1\]

- Support streaming data and queries on them.
- However, they can only handle point locations naively, and do not have adequate supports for evolving spatio-temporal extents.

### Sensor Databases

- Focus more on energy consumption and hiding the inherent heterogeneity and unreliability of sensor networks.
Previous Work

Data Stream Management System (DSMS)[4, 13, 3, 2, 1]
- Support streaming data and queries on them.
- However, they can only handle point locations naively, and do not have adequate supports for evolving spatio-temporal extents.

Sensor Databases
- Focus more on energy consumption and hiding the inherent heterogeneity and unreliability of sensor networks.
In summary, current database systems are NOT capable of handling continuous queries involving both geo-streams and static extended geo-spatial objects.
Streaming Data Types

- **BASE**: int, real, string, bool
- **SPATIAL**: point, points, line, region
- **TIME**: instant
- **WINDOW**: now, unbounded, past
- **RANGE**: range
- **TEMPORAL**: intime, moving
- **STREAM**: streaming

We use $A_\alpha$ to denote the abstract semantics of data type $\alpha$.

**Example**

$A_{points} \equiv \{ P \subseteq \mathbb{R}^2 | P \text{ is finite} \}$
**Window Types**

**now Window**

\[ A_{\text{now}} \equiv \text{current system time} \]

**unbounded Window**

\[ A_{\text{unbounded}} \equiv (-\infty, \text{now}] \]

**past Window**

A time interval proceeding now

\[ A_{\text{past}} \equiv \{ X \subseteq A_{\text{instant}} | \forall x \in X (x \leq \text{now}) \land \forall y \in A_{\text{instant}} (x < y \leq \text{now} \Rightarrow y \in X) \} \]
Window Types

**now Window**

\[ A_{\text{now}} \equiv \text{current system time} \]

**unbounded Window**

\[ A_{\text{unbounded}} \equiv (-\infty, now] \]

**past Window**

A time interval proceeding now

\[ A_{\text{past}} \equiv \{X \subseteq A_{\text{instant}} \mid \forall x \in X (x \leq now) \land \forall y \in A_{\text{instant}} (x < y \leq now \Rightarrow y \in X)\} \]
**Window Types**

```latex
\text{now Window}
\[A_{\text{now}} \equiv \text{current system time}\]

\text{unbounded Window}
\[A_{\text{unbounded}} \equiv (-\infty, \text{now}]\]

\text{past Window}
\[A_{\text{past}} \equiv \{X \subseteq A_{\text{instant}} | \forall x \in X(x \leq \text{now}) \land \forall y \in A_{\text{instant}}(x < y \leq \text{now} \Rightarrow y \in X)\}\]
```
Stream Types

**stream Mapping**

\[ A_{streaming}(\alpha, \omega) \equiv \{ f^c \mid f^c : A_{\text{instant}} \rightarrow A_{\alpha} \} \]

- \( \alpha \): BASE, SPATIAL, or RANGE data type
- \( \omega \): WINDOW data type
- \( f^c \): a partial function that
  1. is undefined for instants not in the window specified by \( \omega \);
  2. is continuously updated over time as the window changes according to the semantics of window type \( \omega \)
Stream Types

Example

\[ A_{\text{streaming}(\text{point}, \text{now})} \equiv \{ f^c | f^c : A_{\text{now}} \rightarrow A_{\text{point}} \text{ where } \forall t \neq \text{now}, f^c \text{ is undefined} \} \]

\[ A_{\text{streaming}(\text{point}, \text{unbounded})} \equiv \{ f^c | f^c : A_{\text{instant}} \rightarrow A_{\text{point}} \text{ where } \forall t \in (\text{now}, \infty), f^c \text{ is undefined} \} \]

- We prefix BASIC or SPATIAL type with an “s” to denote streaming type

Example

`sint`, `sbool`, `spoint`, `sregion`, `srangle`
Stream Types

Example

\[
A_{\text{streaming}(\text{point}, \text{now})} \equiv \{ f^c | f^c : A_{\text{now}} \to A_{\text{point}} \text{ where } \forall t \neq \text{now}, f^c \text{ is undefined} \}
\]

\[
A_{\text{streaming}(\text{point}, \text{unbounded})} \equiv \{ f^c | f^c : A_{\text{instant}} \to A_{\text{point}} \text{ where } \forall t \in (\text{now}, \infty), f^c \text{ is undefined} \}
\]

- We prefix BASIC or SPATIAL type with an “s” to denote streaming type

Example

\text{sint, sbool, spoint, sregion, srange}
Outline

1 Motivation
   - Introduction to Geo-stream
   - Motivation Example

2 Previous Work
   - Previous Work

3 Our Contributions
   - Streaming Data Types
   - Geo-stream Operations
   - Language Embeddings
Windowing Operation

Mapping one stream of window $\omega$ to a stream of window $\psi$

**windowing:** $\text{streaming}(\alpha, \omega) \times \psi \rightarrow \text{streaming}(\alpha, \psi)$

**Example**

$\text{streaming}(\text{point, unbounded}) \times \text{now} \rightarrow \text{streaming}(\text{point, now})$

- We use the notion of $\text{streaming}(\alpha, \omega)[\psi]$ in the language embedding.
Projection to Domain and Range

In Spatio-temporal Database

TEMPPORAL data type → domain and range

In Geo-stream Database

STREAM data type → domain and range represented as another STREAM data type

- Projection is continuously re-evaluated over the moving windows
- The result is of streaming data type
Projection to Domain and Range

Example: Caribous Tracking

```
caribou (name string, location spoint)
```

deftime(caribou.location[unbounded]) returns the times when the caribou is tracked

```
trajectory(caribou.location[past 2 hours]) returns the trajectories (represented as lines) of the caribous in the past two hours continuously
```
### Intersection with Points and Point Sets in Domain and Range

#### Interaction of Streaming Values with Values in Domain and Range

<table>
<thead>
<tr>
<th>Operations</th>
<th>Signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>at instant</td>
<td>$streaming(\alpha, \omega) \times instant \rightarrow intime(\alpha)$</td>
</tr>
<tr>
<td>at periods</td>
<td>$streaming(\alpha, \omega) \times periods \rightarrow moving(\alpha)$</td>
</tr>
<tr>
<td>present</td>
<td>$streaming(\alpha, \omega) \rightarrow streaming(bool, \omega)$</td>
</tr>
<tr>
<td>at</td>
<td>$streaming(\alpha, \omega) \times \alpha \rightarrow streaming(\alpha, \omega)$</td>
</tr>
<tr>
<td>at passes</td>
<td>$streaming(\alpha, \omega) \times range(\alpha) \rightarrow streaming(\alpha, \omega)$</td>
</tr>
<tr>
<td></td>
<td>$streaming(\alpha, \omega) \times \beta \rightarrow streaming(bool, now)$</td>
</tr>
</tbody>
</table>
Example: Caribous Tracking

caribou (name string, location spoint )

city (name string, location point)

```
present(caribou.location[past 3 days])
returns continuously at what times in the past three days the caribous are tracked
passes(caribou.location[past 7 days],city.location)
returns continuously the caribous who had passed a city in the past 7 days
```
Lifting Operations to Streaming Operations

Intersect Example 1

\[ \text{streaming}(\text{region}, \omega) \times \text{region} \rightarrow \text{streaming}(\text{bool}, \omega) \]

- Check if a streaming region intersects with a static region, and produce streaming \textit{bool}.

Intersect Example 2

\[ \text{region} \times \text{streaming}(\text{region}, \omega) \rightarrow \text{streaming}(\text{bool}, \omega) \]

- The semantics is the same as \textit{Intersect} is symmetric.

Intersect Example 3

\[ \text{streaming}(\text{region}, \omega) \times \text{streaming}(\text{region}, \omega) \rightarrow \text{streaming}(\text{bool}, \omega) \]

- Returns a streaming bool of window \( \omega \) representing the intersect results of two streaming regions
Lifting Operations to Streaming Operations

Intersect Example 1

\[
\text{streaming}(\text{region}, \omega) \times \text{region} \rightarrow \text{streaming}(\text{bool}, \omega)
\]

- Check if a streaming region intersects with a static region, and produce streaming \textit{bool}.

Intersect Example 2

\[
\text{region} \times \text{streaming}(\text{region}, \omega) \rightarrow \text{streaming}(\text{bool}, \omega)
\]

- The semantics is the same as \textit{Intersect} is symmetric.

Intersect Example 3

\[
\text{streaming}(\text{region}, \omega) \times \text{streaming}(\text{region}, \omega) \rightarrow \text{streaming}(\text{bool}, \omega)
\]

- Returns a streaming bool of window \(\omega\) representing the intersect results of two streaming regions.
Lifting Operations to Streaming Operations

**Intersect Example 1**

\[ \text{streaming}(\text{region}, \omega) \times \text{region} \rightarrow \text{streaming}(\text{bool}, \omega) \]

- Check if a streaming region intersects with a static region, and produce streaming \text{bool}.

**Intersect Example 2**

\[ \text{region} \times \text{streaming}(\text{region}, \omega) \rightarrow \text{streaming}(\text{bool}, \omega) \]

- The semantics is the same as \text{Intersect} is symmetric.

**Intersect Example 3**

\[ \text{streaming}(\text{region}, \omega) \times \text{streaming}(\text{region}, \omega) \rightarrow \text{streaming}(\text{bool}, \omega) \]

- Returns a streaming bool of window \( \omega \) representing the intersect results of two streaming regions.
Outline

1 Motivation
   - Introduction to Geo-stream
   - Motivation Example

2 Previous Work
   - Previous Work

3 Our Contributions
   - Streaming Data Types
   - Geo-stream Operations
   - Language Embeddings
Language Embeddings

Syntax

SELECT STREAM operation, ...
FROM table with streaming attribute, ...
WHERE STREAM predicate, ...

Motivation Example

forestFire(firename: string, evaArea: sregion)
flood(floodname: string, extent: sregion)
house(owner: string, location: point)
road(roadname: string, extent: line)
Language Embeddings

Syntax

SELECT STREAM operation, ...
FROM table with streaming attribute, ...
WHERE STREAM predicate, ...

Motivation Example

forestFire(firename: \textit{string}, evaArea: \textit{sregion})
flood(floodname: \textit{string}, extent: \textit{sregion})
house(owner: \textit{string}, location: \textit{point})
road(roadname: \textit{string}, extent: \textit{line})
Q1: Notify me when my house is within 50 miles of the mandatory evacuation area of a forest fire.

Answer

```sql
SELECT h.address
FROM house h, forestFire ff
WHERE distance(h.location, ff.evaArea [now]) < 50
```
Motivation Example - Q1

Q1: Notify me when my house is within 50 miles of the mandatory evacuation area of a forest fire.

Answer

SELECT h.address
FROM house h, forestFire ff
WHERE distance(h.location,ff.evaArea [now])<50
Motivation Example - Q2

Q2: Continuously list the addresses of all the houses traversed by flood in the past 2 days in Denton county.

Answer

```sql
SELECT h.address
FROM house h, flood f
WHERE inside(h.location, traversed(f.extent [past 2 days]))
```
Q2: Continuously list the addresses of all the houses traversed by flood in the past 2 days in Denton county.

Answer

```sql
SELECT h.address
FROM house h, flood f
WHERE inside(h.location, traversed(f.extent [past 2 days]))
```
Q3: Continuously list road segments that have been completely under flood-water for the past 24 hours.

Answer

SELECT r.name
FROM road r, flood f
WHERE inside(r.extent,f.extent [past 1 day])
Motivation Example - Q3

Q3: Continuously list road segments that have been completely under flood-water for the past 24 hours.

Answer

SELECT r.name
FROM road r, flood f
WHERE inside(r.extent,f.extent [past 1 day])
Our contributions

1. We propose new streaming and window data types.
2. We investigate the semantics of common spatio-temporal predicates and operations.
3. We illustrate the language embeddings of new proposed data types, predicates and operations.

Future Work

Investigate the discrete representations for each new stream data type
Efficient data models, structures and query optimization algorithms
For Further Reading I


Sirish Chandrasekaran, Owen Cooper, Amol Deshpande, Michael J. Franklin, Joseph M. Hellerstein, Wei Hong, Sailesh Krishnamurthy, Samuel R. Madden, Fred Reiss, and Mehul A. Shah.

Telegraphcq: continuous dataflow processing.

Jianjun Chen, David J. DeWitt, Feng Tian, and Yuan Wang.
Niagaracq: a scalable continuous query system for internet databases.
*SIGMOD Rec.*, 29(2), 2000.
For Further Reading III


Ralf Hartmut Güting, Michael H. Böhlen, Martin Erwig, Christian S. Jensen, Nikos A. Lorentzos, Markus Schneider, and Michalis Vazirgiannis.

Ralf Hartmut Güting, Victor Teixeira de Almeida, Dirk Ansorge, Thomas Behr, Zhiming Ding, Thomas Höse, Frank Hoffmann, Markus Spiekermann, and Ulrich Telle.
Secondo: An extensible dbms platform for research prototyping and teaching.
For Further Reading V


For Further Reading VI

Lukas Relly and Uwe Röhm.  
Plug and play: Interoperability in concert. 

Peter A. Tucker, David Maier, Tim Sheard, and Leonidas Fegaras.  
Exploiting punctuation semantics in continuous data streams.  