GridVine: Building Internet-Scale Semantic Overlay Networks

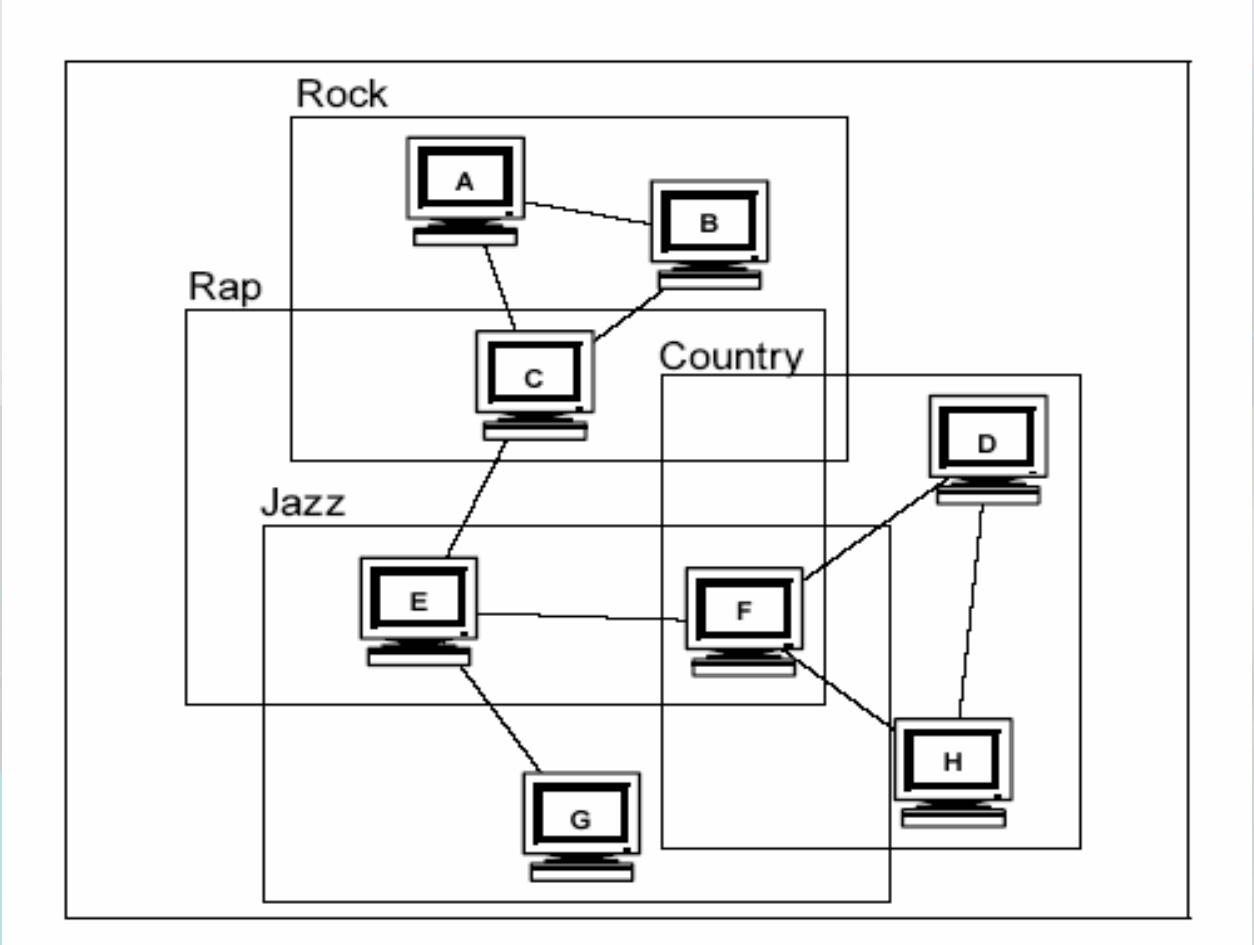
By Lan Tian

Outline

- Introduction
- Overview of our Approach
- The P-Grid P2P system
- Semantic Support
- Semantic Interoperability
- Related Work
- Summary

Introduction

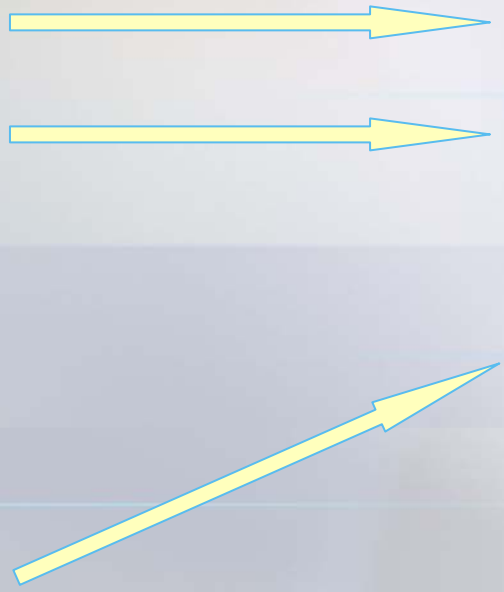
What is semantic Overlay Network?



Schema mapping

- ARTICLE
  - artPK
  - title
  - pages
- AUTHOR
  - artFK
  - name

- PUBLICATION
  - pubID
  - title
  - date
  - author



```
SELECT artPK AS pubID        UNION SELECT   null AS pubID
       title AS title                       null AS title
       null AS date                         null AS date
       null AS author                       name AS author
FROM   ARTICLE                     FROM    AUTHOR
```

Introduction

- Based on federated databases
- Realization of semantic overlay networks in order to enable semantic interoperability
- Key aspect: data independence
  - logical layer:

    supports semantic interoperability;

    provides semantic gossiping
  - physical layer:

    provides operations exploiting a structured overlay network,P-Grid

Introduction

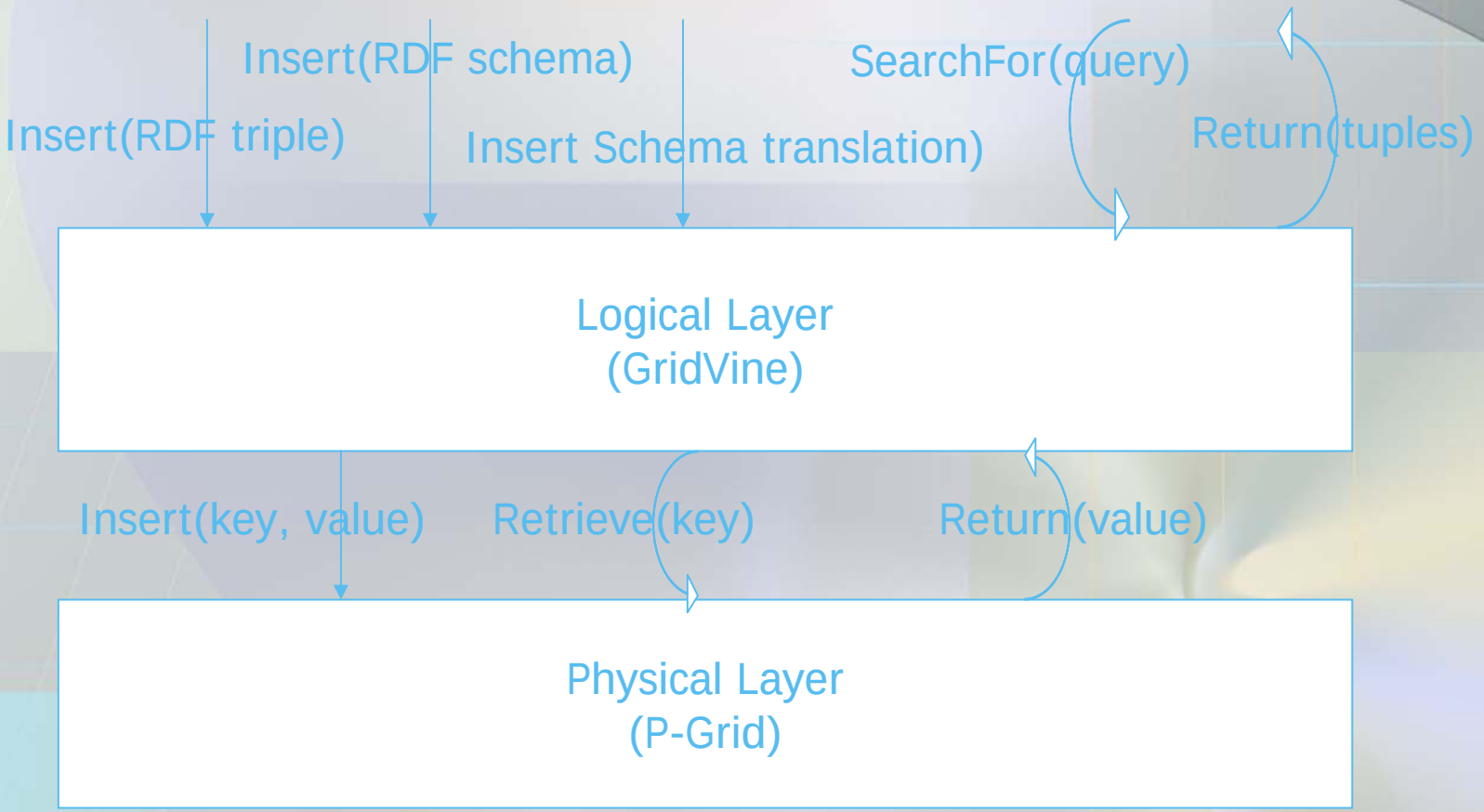
- Requires mappings of operations and data to the physical layer
  - Introduction of a specific name space present in the peer space
  - The mapping of data and metadata to routable keys
  - The implementation of traversals of the semantic network for querying using intermediate schema mapping

Outline

- Introduction
- **Overview of our Approach**
- The P-Grid P2P system
- Semantic Support
- Semantic Interoperability
- Related Work
- Summary

Data Independence

- **Two-layer model**



Data Independence

- P-Grid is an efficient, self-organizing and fully decentralized access structure
- GridVine uses two of P-Grid‘s basic functionlities: Insert(key, value) and Retrieve(key)
- Choose RDF/RDFS as languages to encode metadata and vocabulary definitions in GridVine

Decentralized Semantics

- Schema inheritance

  - provides GridVine with basic schema reusability and interoperability

- Semantic Gossiping

  - applied to foster semantic interoperability in decentralized settings

  - Query forwarding

    - iterative forwarding
    - recursive forwarding

Outline

- Introduction
- Overview of our Approach
- **The P-Grid P2P system**
- Semantic Support
- Semantic Interoperability
- Related Work
- Summary

The P-Grid P2P system

- GridVine uses P-Grid system as physical layer
- P-Grid is based on the principles of distributed hash tables(DHT)
- Peers refer to a binary tree structure
- Each peer p $\in$ P is associated with a leaf
- Each leaf corresponds to a binary string $\Pi \in \prod$

  $\dashrightarrow$ each peer associated with a path
- Each peer stores a set of data items δ (p)
- For d $\in$ δ (p) binary key key(d) has Π (p) as prefix

Outline

- Introduction
- Overview of our Approach
- The P-Grid P2P system
- **Semantic Support**
- Semantic Interoperability
- Related Work
- Summary

Semantic Support

- **Metadata Storage**
  - **URI Schemes p-grid :/ /, for resource, p-grids:/ /, for schema-elements**
  - **all resources identified by P-Grid URIs**
  - **Example :**

P-Grid resource 11110101(subject) is entitled(predicate) Rain, Stream and Speed(object)

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
    <rdf:Description rdf:about="pgrid://11110101">
        <Title xmlns="pgrids://01001101:bmp#">Rain, Steam and Speed</Title>
    </rdf:Description>
</rdf:RDF>
```

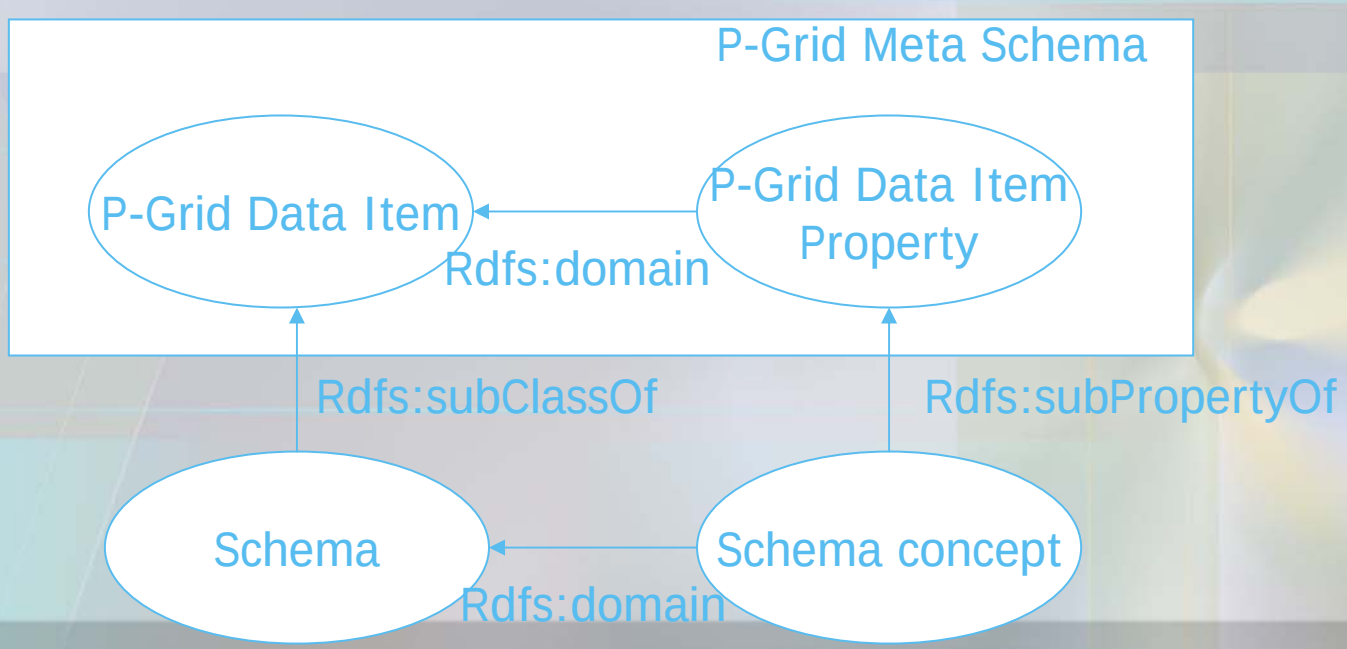
Metadata Storage

- We reference each individual triple three times, generating seperate keys based on their subject, predicate and object values
- Insertion operation of a triple $t \in T$

$$\begin{aligned} Insert(t) = \; & Insert(t_{subject}, t), \\ & Insert(Hash(t_{predicate}), t), \\ & Insert(Hash(t_{object}), t) \end{aligned}$$

Schema Definition And Storage

- Schematic information in GridVine is encoded using RDFS
- P-Grid meta-schema and its relation to user-defined RDF schemas



Resolving Queries in GridVine

- Simplest query with one bound variable

  SELECT ?y

  WHERE (<p-grid://01101000>, ?y, ?z)

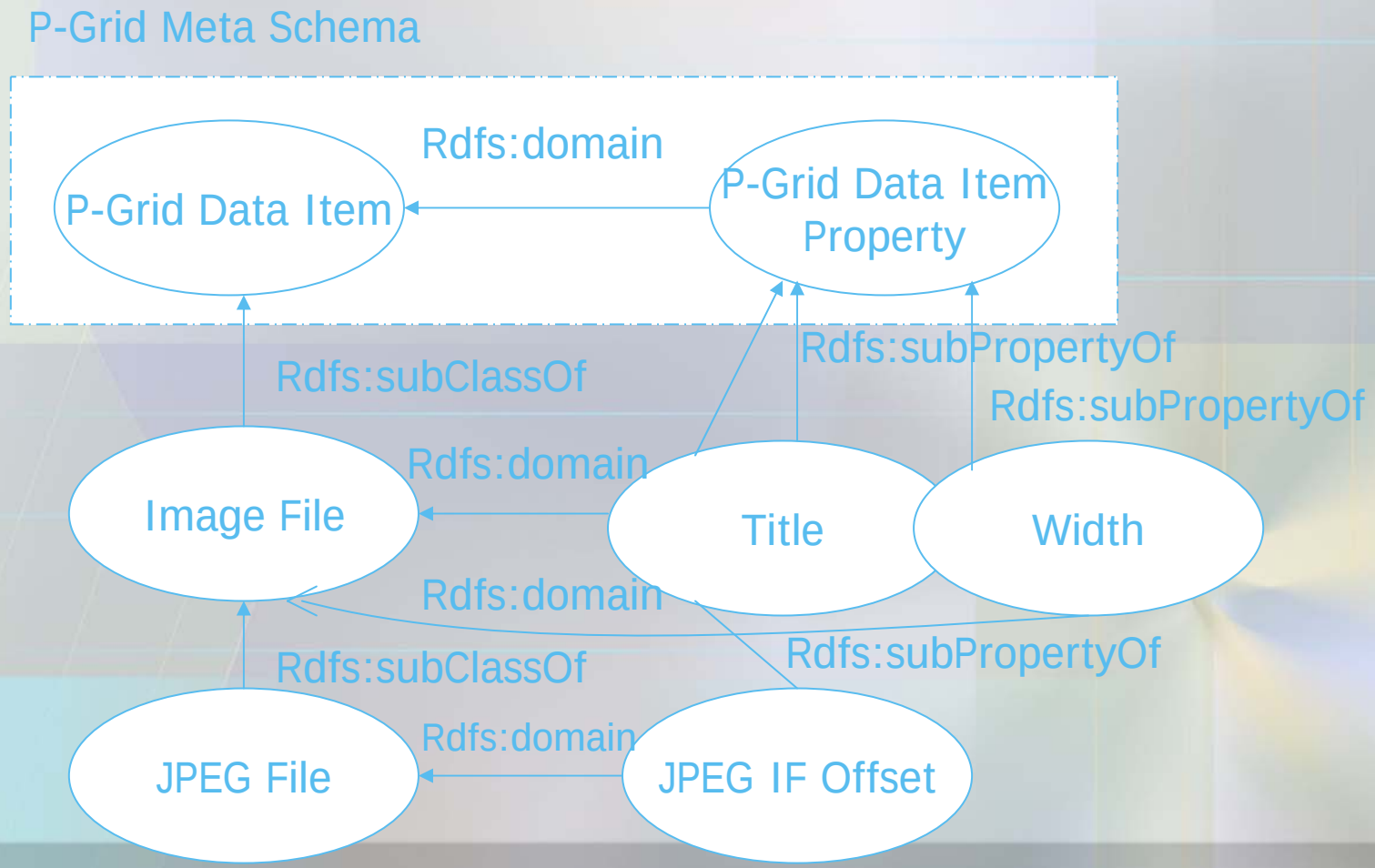
  - return all the predicates used to annotate data item 01101000
- Implementation
- Subject ,predicate and object can all be replaced by variables which may be bound or not

Outline

- Introduction
- Overview of our Approach
- The P-Grid P2P system
- Semantic Support
- **Semantic Interoperability**
- Related Work
- Summary

Semantic Interoperability

- **Schema Inheritance**



Semantic Gossiping

- Aims at establishing global forms of agreement starting from a graph of purely local mappings among schemas
- Semantic neighbourhood
- Network can be seen as a directed graph of translations
- Two interesting properties :

  - Transitivity allows for the forwarding of queries to semantic domains for which there is no direct translation link

  - Check the quality of translations

Semantic Gossiping

- **Translation links are stored as OWL documents**

```
<?xml version="1.0"?> <?xml version="1.0"
encoding="ISO-8859-1" ?>
<rdf:RDF xmlns:owl ="http://www.w3.org/2002/07/owl#"
xmlns:rdf ="http://www.w3.org/1999/02/22-rdf-syntax-
ns#">
<Image_Description xmlns="pgrids://10000101:exif#">
<owl:equivalentProperty rdf:ID="m1"
rdf:resource="pgrids://01001101:bmp#Title"/>
</Image_Description>
<Exif_Image_Width xmlns="pgrids://10000101:exif#">
<owl:equivalentProperty rdf:ID="m2"
rdf:resource="pgrids://01001101:bmp#Width"/>
</Exif_Image_Width>
</rdf:RDF>
```

Outline

- Introduction
- Overview of our Approach
- The P-Grid P2P system
- Semantic Support
- Semantic Interoperability
- **Related Work**
- Summary

Related Work

- **Hyperion:**

  proposes an architecture and outlines a set of challenges for decentralized data management in P2P Systems
- **SWAP:**

  Approach combining P2P and Semantic Web techniques
- **Edutella:**

  employs a super-peer topology and facilitates the clustering of data based on ontology, rule, or query

Related Work

- PeerDB:

  - facilitates sharing of data without shared schema;

  - combines the power of mobile agents into P2P systems to perform operations at peers' sites

- Pizza peer data management project:

  - takes an approach to semantic heterogeneity that is similar to Semantic Gossiping;

  - provides no measures to judge correctness

Summary

- Introduction
- Overview of our Approach
  - Data Independence
  - Decentralized Semantics
- The P-Grid P2P system
- Semantic Support
  - Metadata Storage
  - Schema Definition And Storage
  - Resolving Queries in GridVine
- Semantic Interoperability
  - Schema Inheritance
  - Semantic Gossiping
- Related Work