Towards Scalable Grid Ontology Repository

Marian Babik

Dept. of parallel and distributed computing, IISAS

4th Cracow Grid Workshop, 2004

Contents

- **Motivation**
- **Semantic Web Services**
  - IRS-II, WSMF, OWL-S
- **Design**
  - Distributed hash table systems
  - Ontologies
- **Evaluation**
- **Conclusion**

Motivation

- **Automatic Web Service (WS) discovery**
  - Automatic location, constraints
- **Automatic WS invocation (execution)**
  - Declarative APIs, executing collection of function calls
- **Automatic WS composition, interoperation and publication**
  - High-level description of objective (coop w. data mng.)
- **Automatic WS execution monitoring**

Semantic Web Services Frameworks

- **Web Service Modelling Framework (WSMF)**
  - joint effort of EU research projects
  - industry scale framework
  - WSMO
  - WSML
  - WSMX
  - de-coupling of components, languages with different expressive power (descr. logic, horn logic)
  - semantic WS interacting only through mediators

- **Internet Reasoning Service (IRS-II)**
  - Open University, London
  - UPML framework (task model, problem solving methods (PSM), domain model, bridges)
  - OCML (Operational Conceptual Modelling Language) – procedural knowledge
  - can import OWL-S

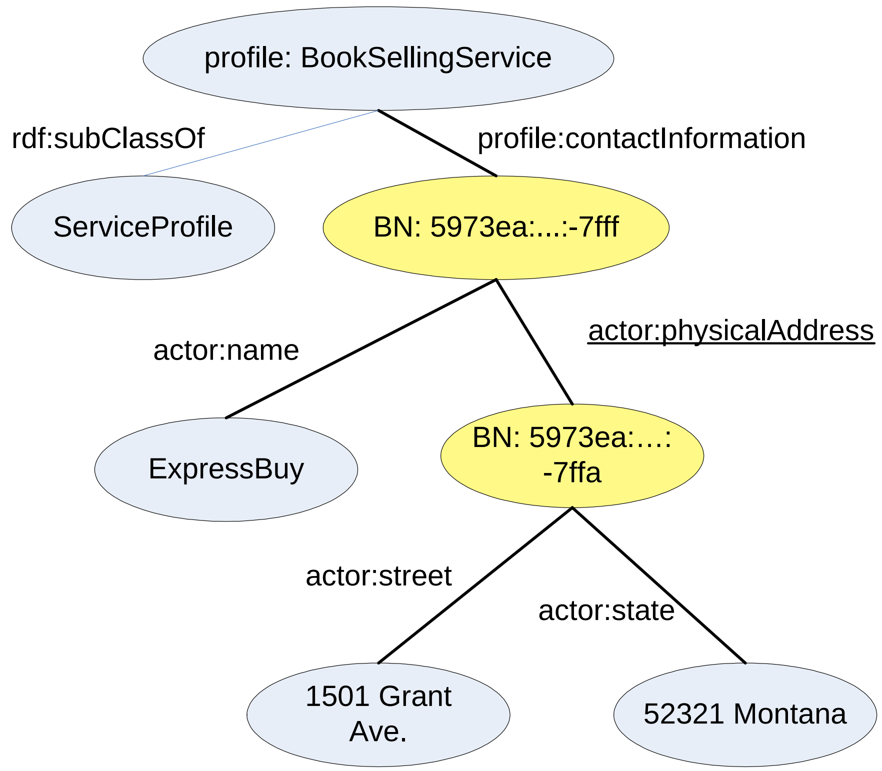
- **OWL-S (DAML-S)**
  - BBN, Carnegie-Mellon, Nokia, Stanford, SRI Int., Yale
  - Upper ontology
  - Service, Profile, Process, Grounding

Scalable Grid Ontology Repository

- Distributed
- Scalable
- Indexing ontologies
- High Availability
- Guaranties to find answers if they exists
- Collaborative environments
- Support automatic WS composition
- Support
  - RDF languages (RDQL, SPARQL)
  - Distributed reasoning (OWL-QL)
- Based on OWL/OWL-S

Ontologies

- **Ontologies** – shared and common understanding of a domain, that can be communicated btw. people and apps.
  - define formal semantics for information
  - define real-world semantics
- **Languages**
  - W3C OWL (RDF, RDFS, XML)
- **W3C OWL Web Ontology Language Semantics and Abstract Syntax**
- **Blank nodes**

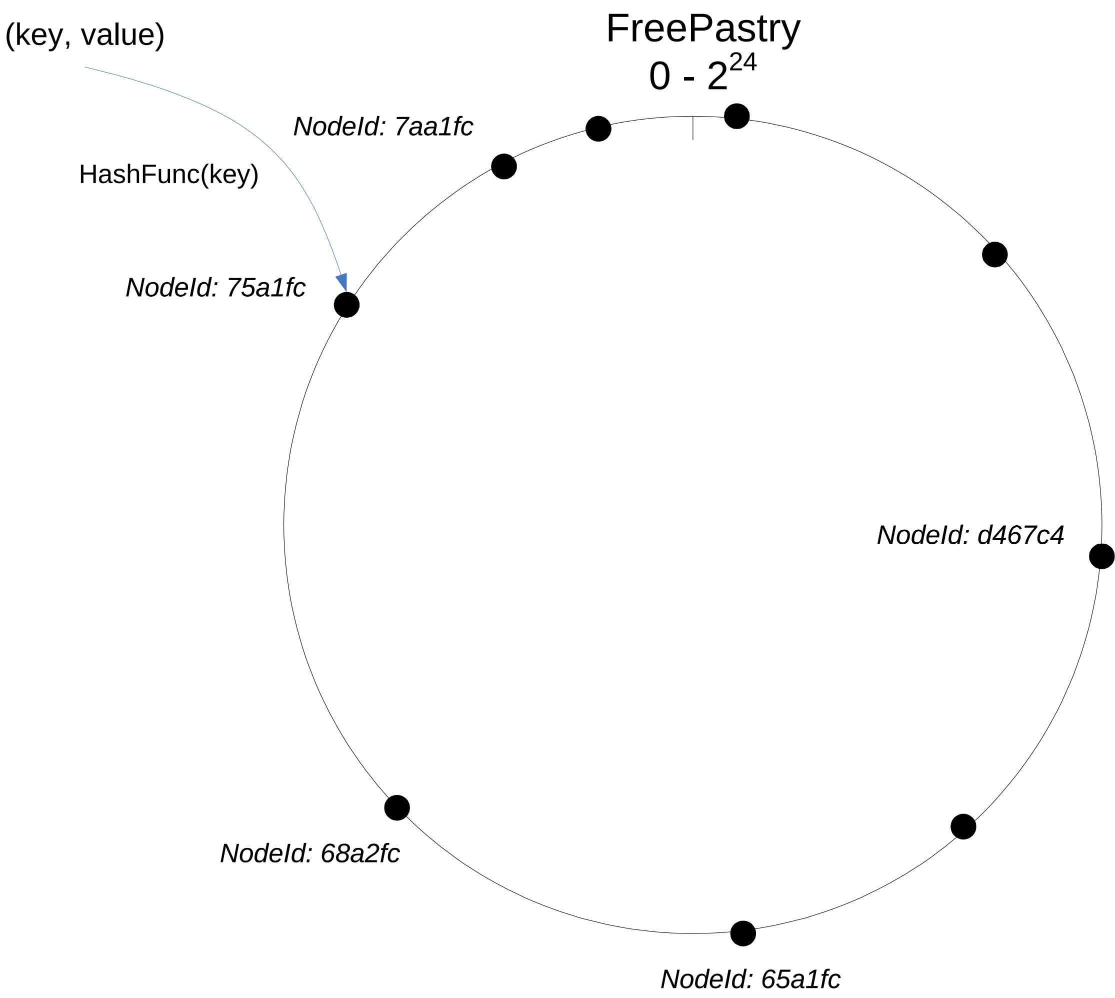


Subject	Predicate	Object
Profile:BookSellingService	rdf:subClassOf	ServiceProfile
Profile:BookSellingService	profile:contactInformation	BN:5973ea:9387:-7fff
BN: 5973ea:9387:-7fff	actor:name	ExpressBuy
BN: 5973ea:9387:-7fff	actor:physicalAddress	BN:5973ea:9993:-7ffa
BN:5973ea:9993:-7ffa	actor:street	“1501 Gran Ave.”
BN:5973ea:9993:-7ffa	actor:state	“52321 Montana”

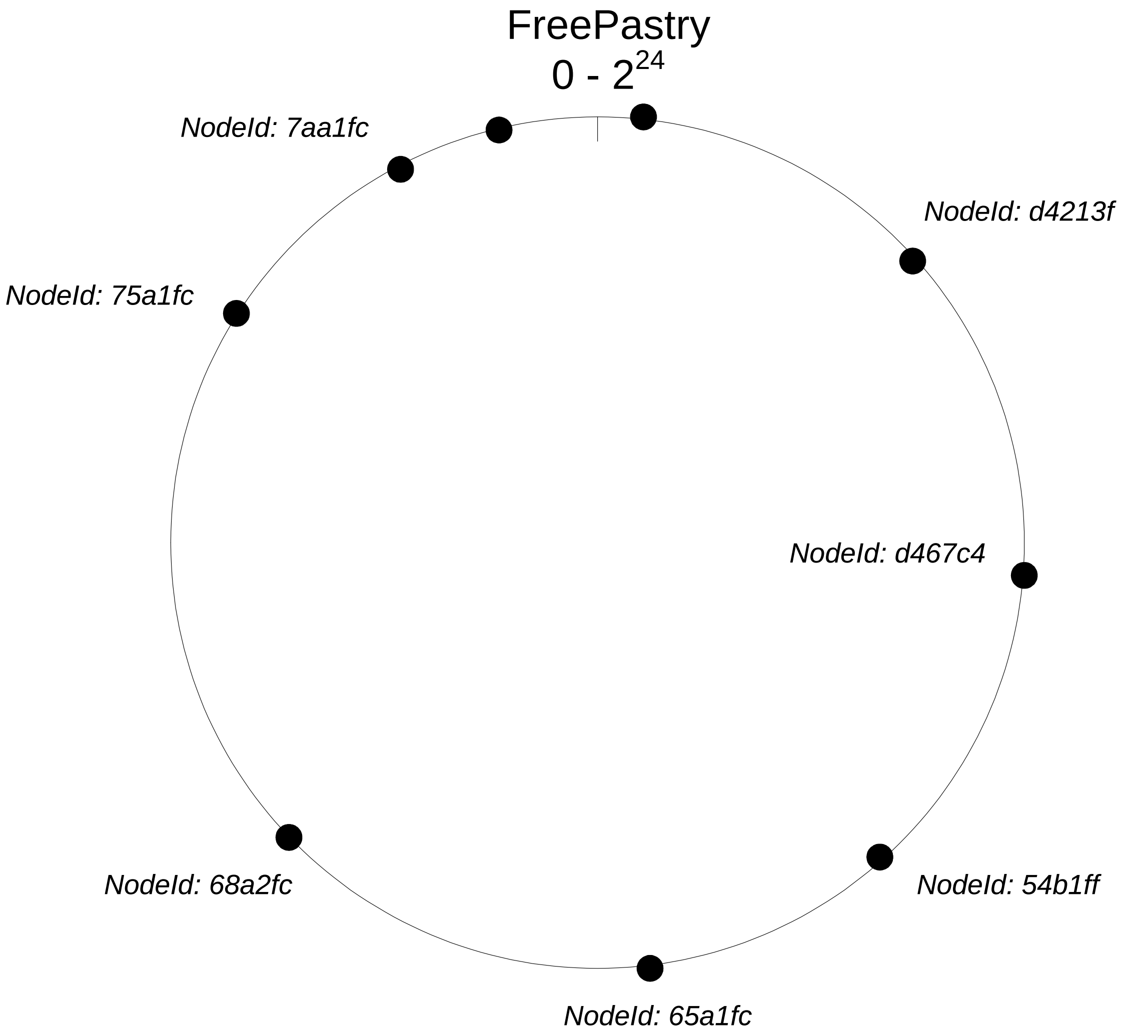
profile: http://www.daml.org/services/owl-s/1.1/Profile.owl
actor: http://www.daml.org/services/owl-s/1.1/ActorDefault.owl
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns
owl: http://www.w3.org/2002/07/owl

Distributed Hash Tables (DHT)

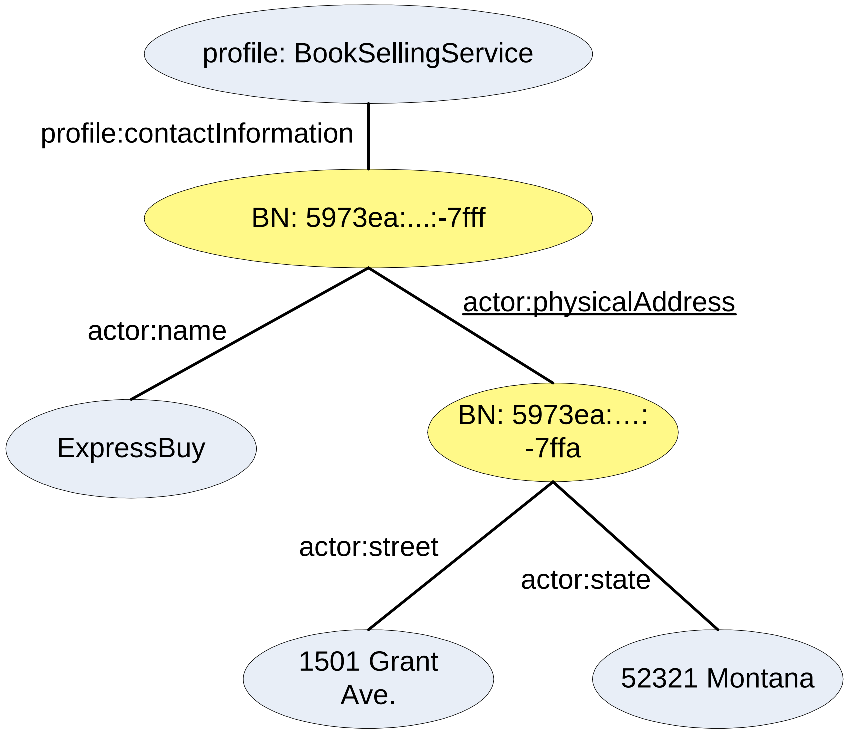
- Distributed
- Scalable
- Self-organizing
- Routing complexity log(N)
- Routing table log(N)
- Kademlia, Chord, CAN, Tapestry, Pastry
- Pastry
  - Network locality
  - Complexity
  - Implementation

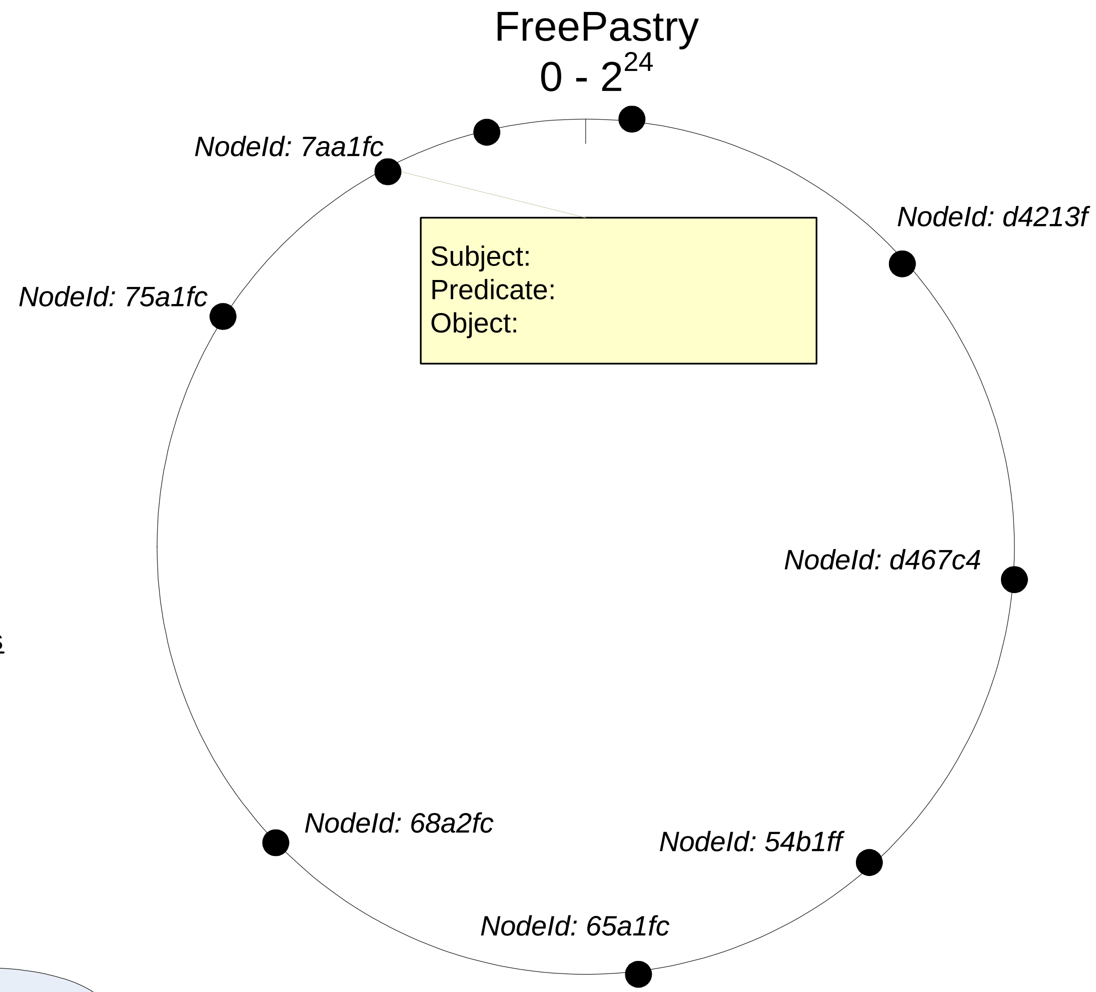


Design


FreePastry
0 - 2^24
NodeId: 7aa1fc
NodeId: d4213f
NodeId: 75a1fc
NodeId: d467c4
NodeId: 68a2fc
NodeId: 54b1ff
NodeId: 65a1fc

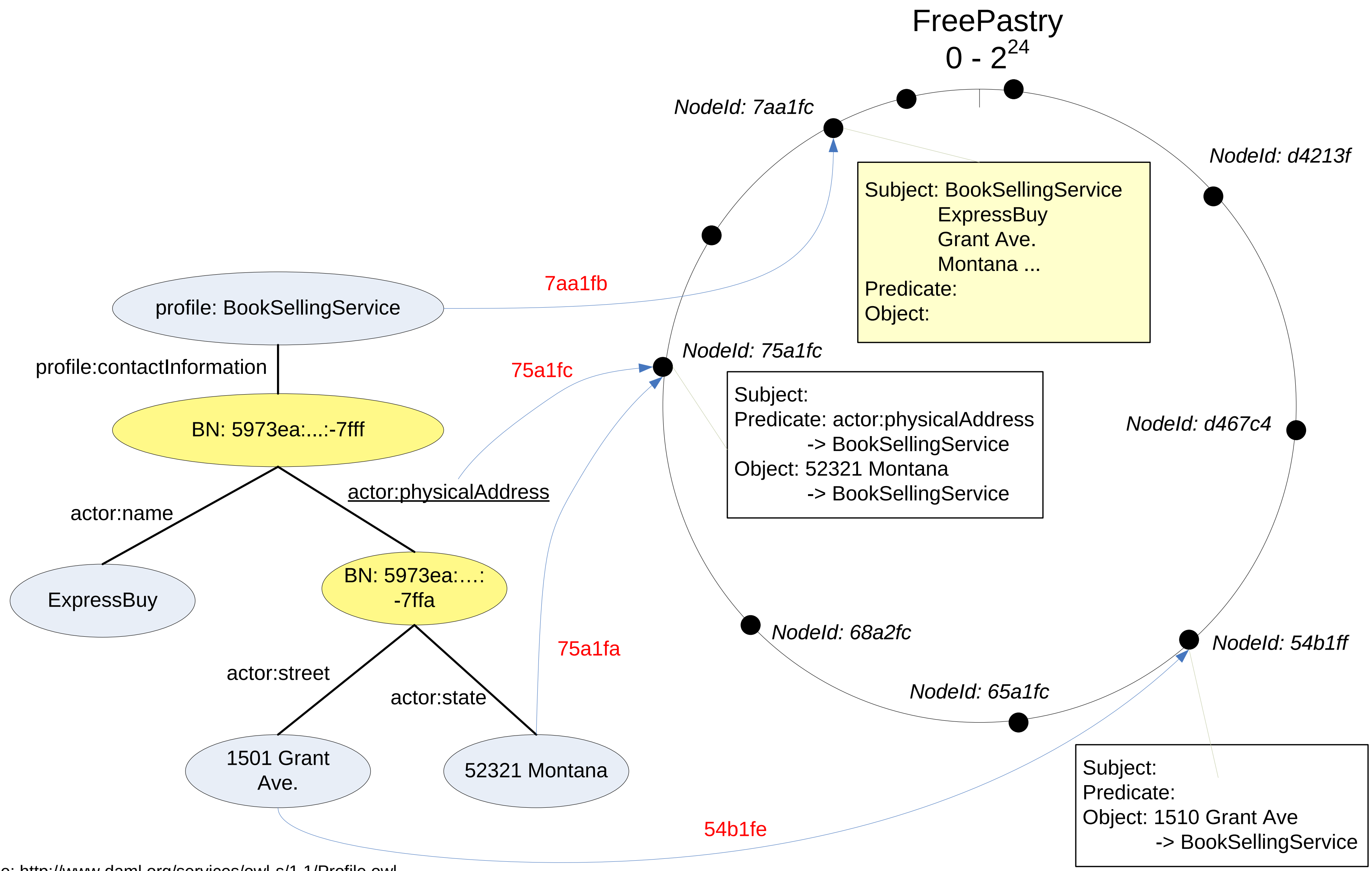
Design (1)





profile: http://www.daml.org/services/owl-s/1.1/Profile.owl
actor: http://www.daml.org/services/owl-s/1.1/ActorDefault.owl

Design (2)

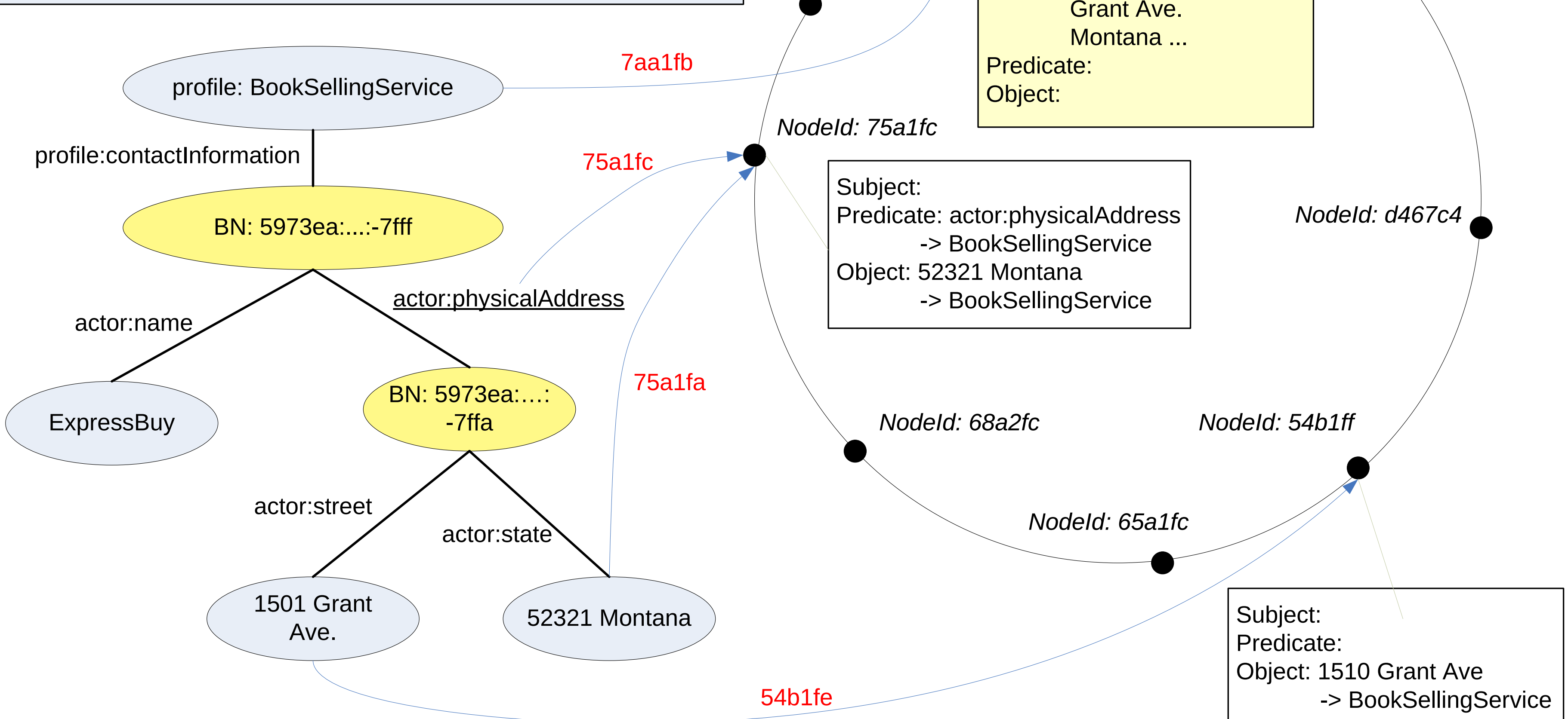


profile: http://www.daml.org/services/owl-s/1.1/Profile.owl
actor: http://www.daml.org/services/owl-s/1.1/ActorDefault.owl

Design (3)

BN – blank node (anonymous node); O – ontology decomposed into triples (rdf:Statements)
Method parse(ontology O)
for every rdf:Statement = <rdf:subject, rdf:predicate: rdf.object> in ontology O
    if rdf:subject is not BN and rdf:object is BN
        resolveBN(rdf:object)

Method resolveBN(rdf:object):
    for every rdf:Statement = <rdf:subject2, rdf:predicate2, rdf:object2> in ontology O
        if rdf:subject2 == rdf:object and rdf:object2 is BN
            resolveBN(rdf:object2)



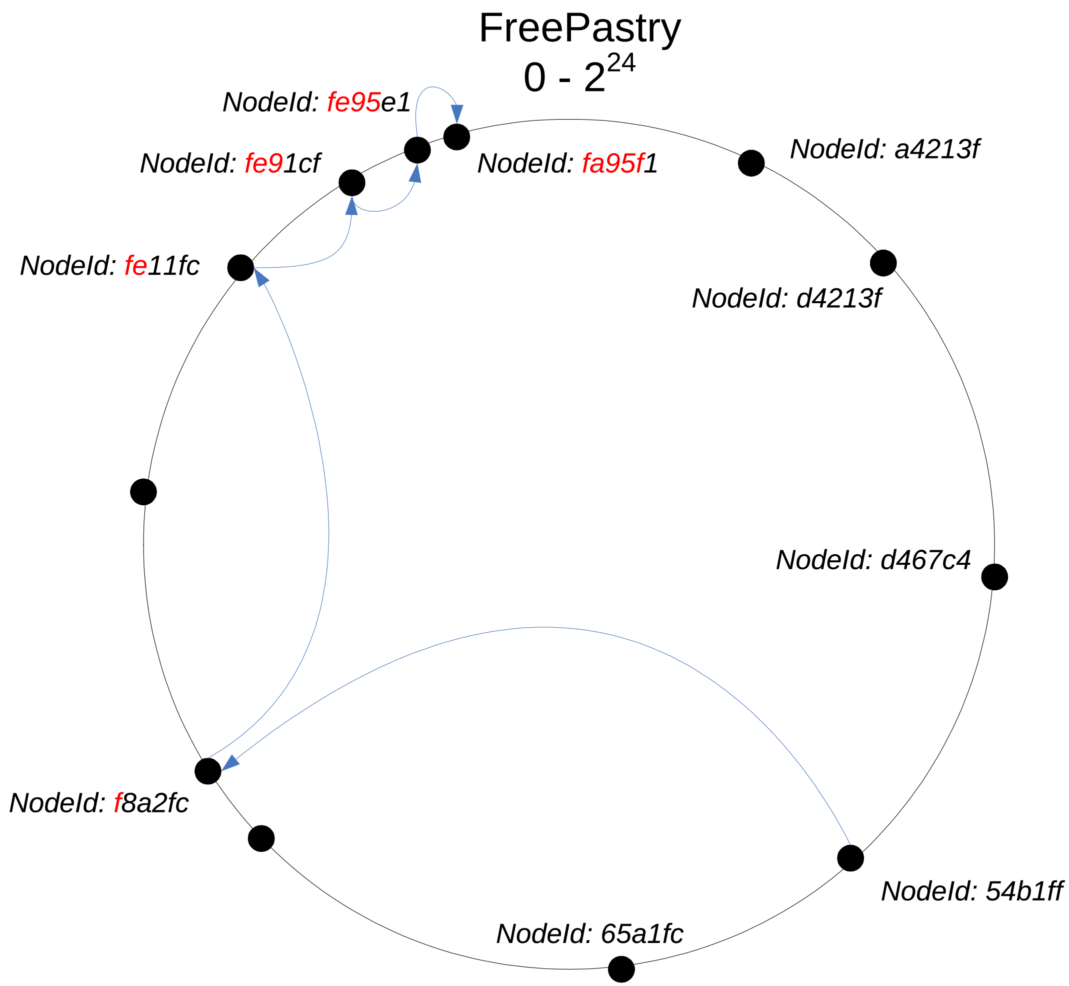
profile: http://www.daml.org/services/owl-s/1.1/Profile.owl
actor: http://www.daml.org/services/owl-s/1.1/ActorDefault.owl

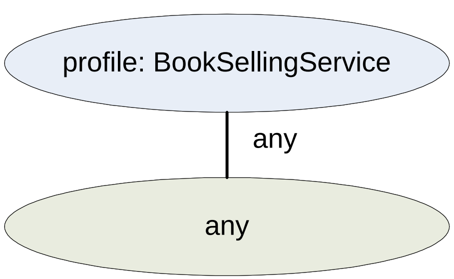
Publishing and querying

For each routing step:

- Node forwards message to a node whose NodeId shares with the key a prefix that is at least one digit longer that the key shares with the present node's id

- If no such node is known, forward message to a node whose NodeId is numerically closer to the key than the present node's id.

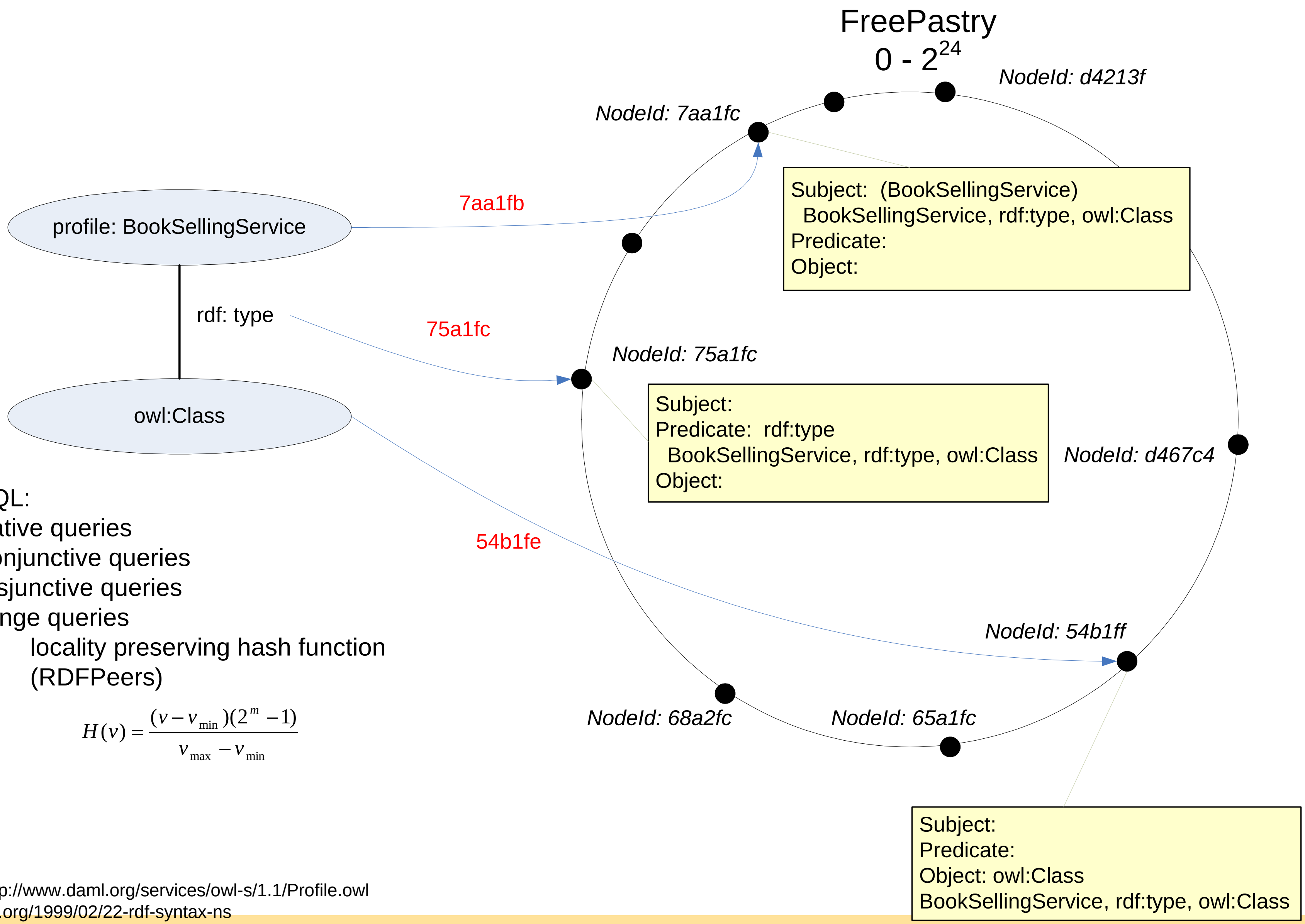




Query: < rdf:subject = profile: BookSellingService; rdf:predicate?, rdf:object?>
At node 54b1ff; hashF(rdf:subject) = fe95f6

profile: http://www.daml.org/services/owl-s/1.1/Profile.owl
actor: http://www.daml.org/services/owl-s/1.1/ActorDefault.owl

Simplified



RDQL:
- native queries
- conjunctive queries
- disjunctive queries
- range queries
  locality preserving hash function (RDFPeers)

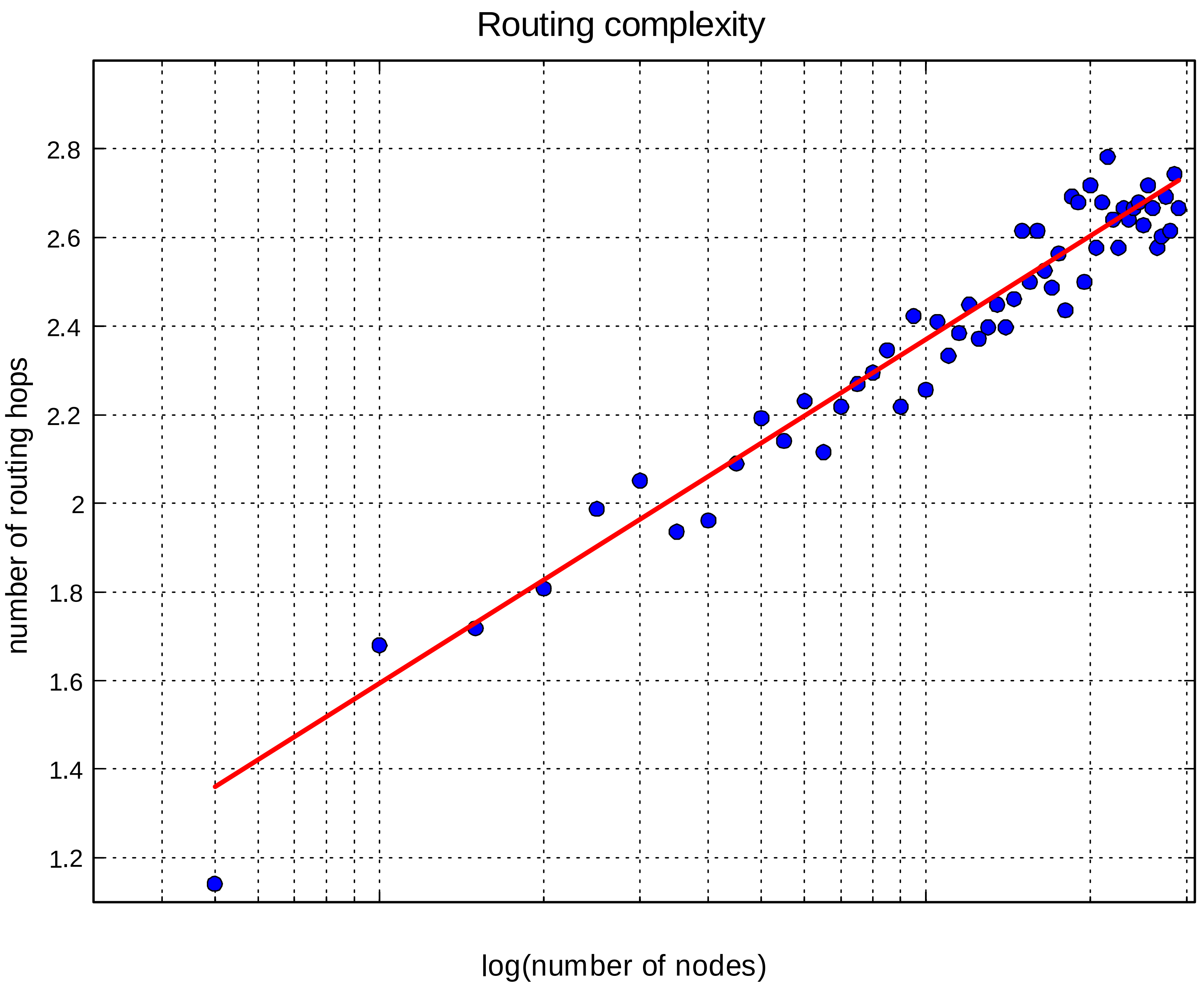
$$H(v)=\frac{(v-v_{\min})(2^{m}-1)}{v_{\max}-v_{\min}}$$

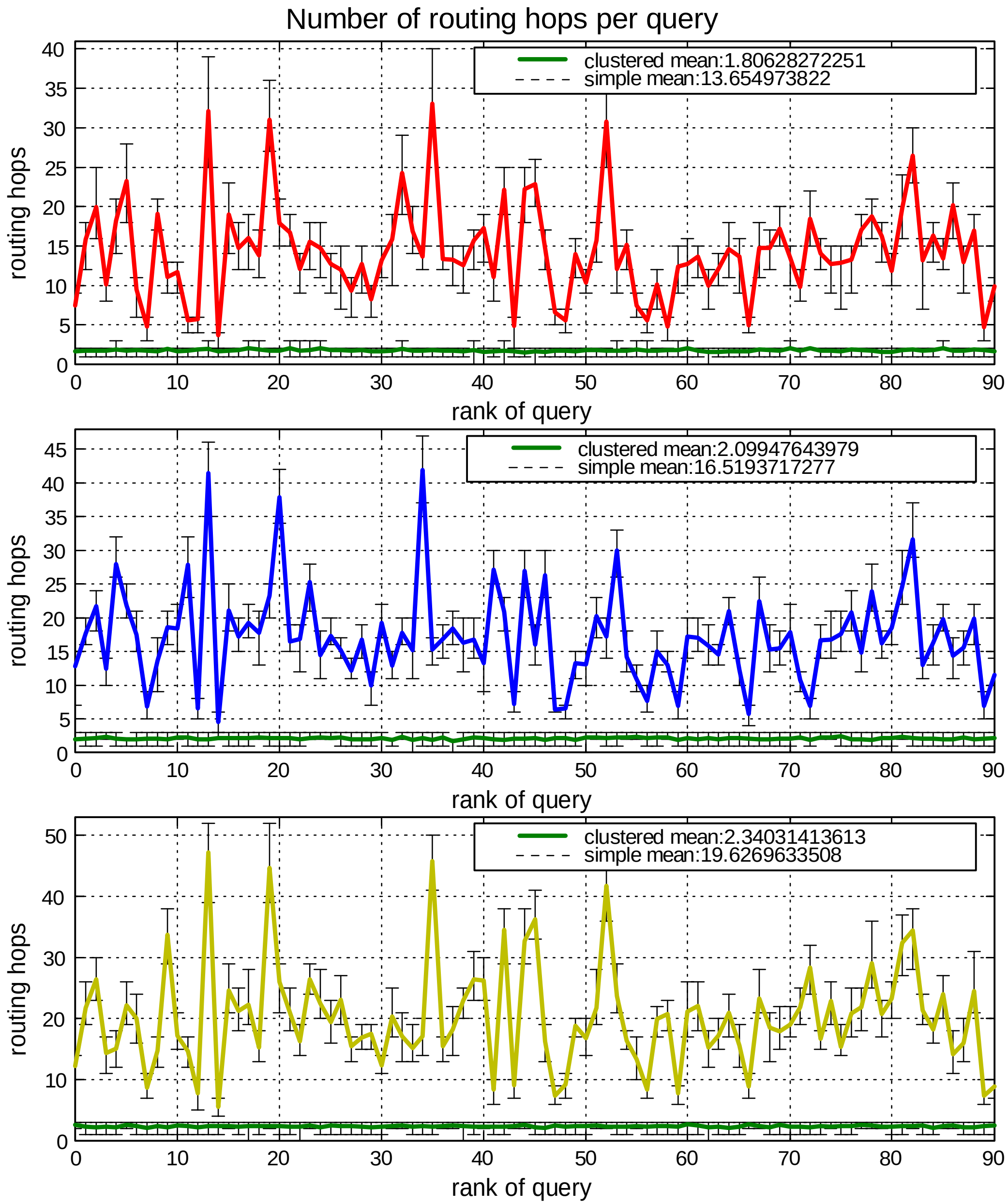
OWL-S profile: http://www.daml.org/services/owl-s/1.1/Profile.owl
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns
owl: http://www.w3.org/2002/07/owl

Evaluation

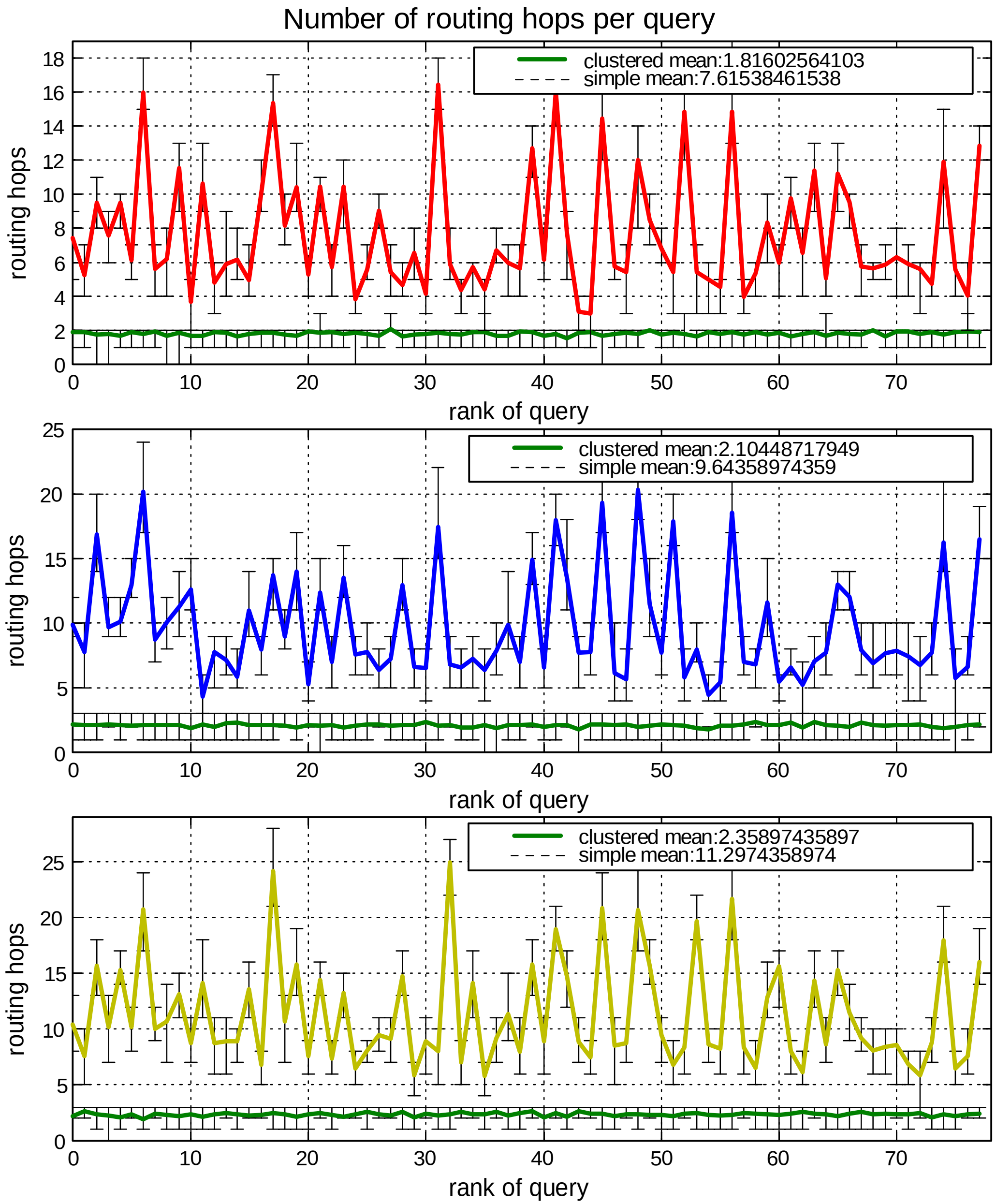
- **Tests**
  - Routing complexity (network hops) – 3000 nodes
  - Comparison of routing complexity btw. simple and clustered model (200,500,1000 nodes)
- **Implementation**
  - Jena – semantic web framework
    - Simulation driver
    - Distributed driver – event-driven architecture
  - FreePastry
    - Open source implementation of Pastry

Ontology	Triples	Blank nodes	Description
BookSellerService	1083	557	Book selling service (OWL-S example)
OWL-S ontologies	2032	869	Owl-S Profile, Grouding, Service, Process etc.
Tambis	4200	3366	Biological science, vocabulary
Iso-19115	5288	1818	Geographical metadata information
Fgdc	3711	1918	Content Standard for digital geospatial metadata
Overall	15231	7991	

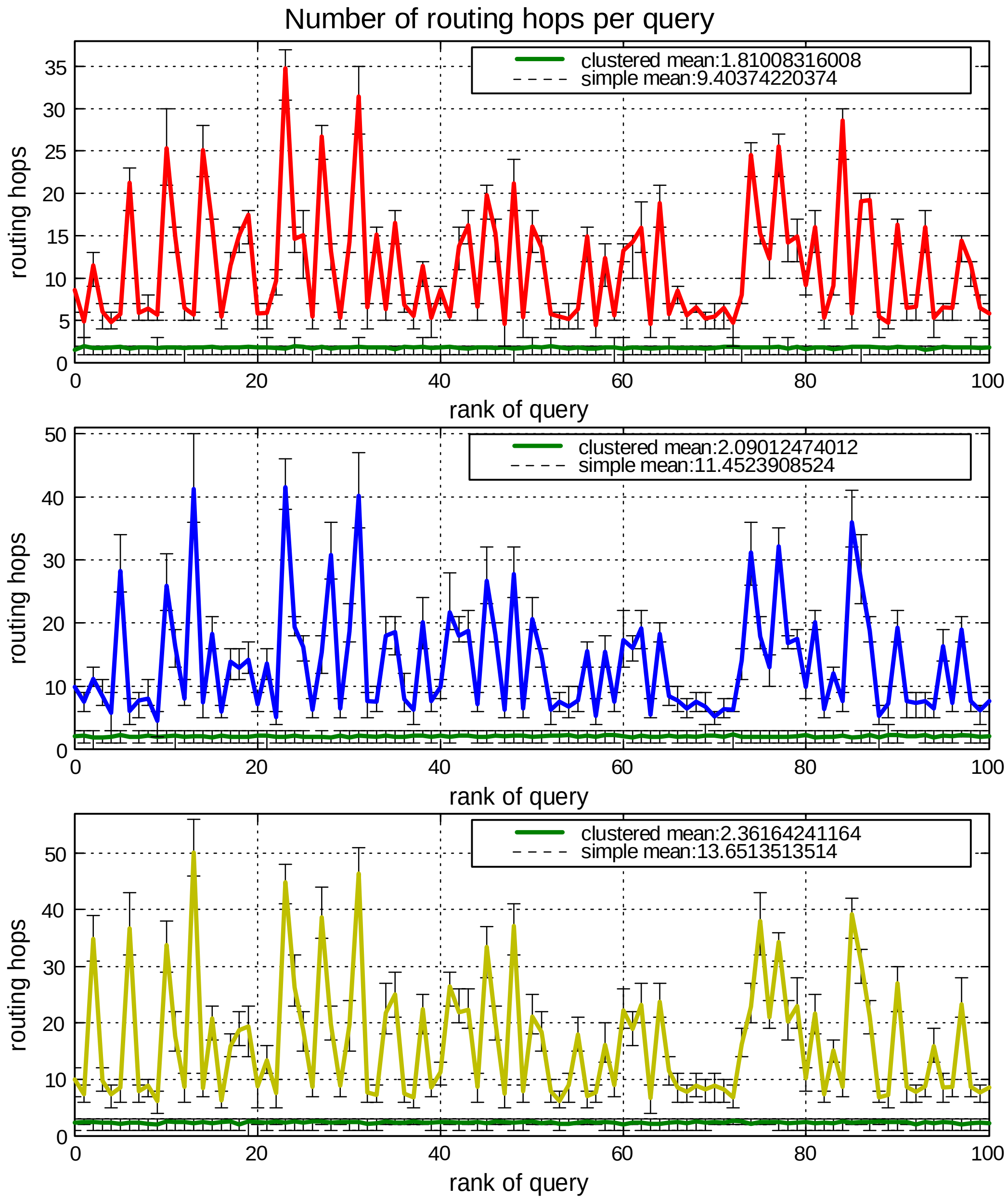

Routing complexity
number of routing hops
log(number of nodes)
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2



- Tambis
- 4200 triples
- 3366 blank nodes
- Biological science, vocabulary
- Experiment:
  - 200 nodes
  - 500 nodes
  - 1000 nodes
- Simple vs. Clustered method



- OWL-S
- 3155 triples
- 1042 blank nodes
- BookSellingService, OWL-S ontologies
- Experiment:
  - 200 nodes
  - 500 nodes
  - 1000 nodes
- Simple vs. Clustered method



- Overall
- 15231 triples
- 7991 blank nodes
- all ontologies
- Experiment:
  - 200 nodes
  - 500 nodes
  - 1000 nodes
- Simple vs. Clustered method

Conclusion

- **Distributed model**
- **Scalability**
- **Indexing**
- **Load balancing**
  - Virtual nodes
  - Simple methods
  - Proximity aware methods
- **RDQL**
- **Distributed reasoning**
  - OWL-QL
- **Performance**

Backup slides

Related work

- RDF-Peers (MAAN)
- OWL-S/DAML-S UDDI
- Piazza
- Edutella
- oceanStrore, PAST etc.