Semantic Interpretation of Internet Monitoring Data

Ian Napier
PhD Student, Dept of Computer Science
Loughborough University

Supervisors : Dr Iain Phillips
Dr Lin Guan

Multi-Service Networks Workshop, Cosener’s House 2009
Example Scenario: Intersite VPN
Where's the gap?
• Different published formats (PCAP files, CSV, HTML, spreadsheets, databases etc.) How to access and combine these?

• Inconsistent or Conflicting data (eg: AS level topology from BGP updates or Active probing methods [4]). How to mediate?

• Understanding the significance of, and relationship between disparate data. Interpretation and semantics need domain expertise. How to encapsulate this knowledge?

• RDF data models and Ontologies are one approach.
A nice general definition:

“An Ontology is a structure capturing semantic knowledge about a certain domain by describing relevant concepts and relations between them.” [5]
Layered Approach, built on the foundation of the Resource Description Framework (RDF) and the RDF ‘triple’

Statement:

\[ \text{subject} \rightarrow \text{predicate (property)} \rightarrow \text{object} \]

cos:Cosener’s Hotel \rightarrow cos:locatedIn \rightarrow cos:Oxfordshire

“149,100” \rightarrow cos:hasPopulation \rightarrow cos:Oxford

cos = “http://nets.lboro.ac.uk/home/coidn/cosenersontology.rdf/”
Data Modelling (RDF)

Ontology Languages (RDFS and OWL)

Reasoners (Pellet, Fact) / Rules (SWRL)

Query (SPARQL)

Semantic Web Languages
Data Converters (eg: PCAP to RDF)

"Local" data (eg: Packet traces, traceroutes)

Data Converters (eg: PCAP to RDF)

"Public" data (eg: RIPE)

Queries

Reasoner/Rules

Ontology

RDF Triple Store

Serialisers/Parsers

RDF/XML files

Application

Multi-Service Networks Workshop, Cosener's House 2009
Sample Ontology: Graph snippet

Multi-Service Networks Workshop, Cosener’s House 2009
Sample Ontology: RDF/XML snippet

```xml
<rdf:Description rdf:about="http://nets.lboro.ac.uk/TCPFLOW72">
  <lboro:hasRESET>false</lboro:hasRESET>
  <lboro:hasSRCIPAddr rdf:resource="http://nets.lboro.ac.uk/SRCIP72"/>
  <lboro:hasDSTPort>80</lboro:hasDSTPort>
  <rdf:type rdf:resource="http://nets.lboro.ac.uk/TCPFlow"/>
</rdf:Description>

<rdf:Description rdf:about="http://nets.lboro.ac.uk/UDPFlow">
  <owl:disjointWith rdf:resource="http://nets.lboro.ac.uk/TCPFlow"/>
  <rdfs:subClassOf rdf:resource="http://nets.lboro.ac.uk/IPCommunication"/>
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
</rdf:Description>
```
Next Steps

• Expand ontology domain to include more data, specifically more converters.
• Develop richer semantics (predicates).
• Develop queries and application, liaising with industry.
• Initially concentrate on AS topology data and local packet captures.

Thank You. Any questions?
References and Reading

References


Recommended Reading


Related Work

http://www.fp7-moment.eu/

Multi-Service Networks Workshop, Cosener’s House 2009
**Supplementary Notes: Some Properties**

- rdf
  - :type
- rdfs
  - :subClassOf
  - :domain
  - :range
  - :subPropertyOf
- owl
  - :class
  - :ObjectProperty
  - :datatypeProperty
  - :inverseOf
  - :SymmetricProperty
  - :ReflexiveProperty $A \ p A$ for all $A$
  - :IrreflexiveProperty
  - :TransitiveProperty
“Locally generated” data, including Passive packet captures at Internet gateways, Active traceroute data, etc.

Regional Internet Registry data (eg: RIPE): delay measurements, AS to IP Address mappings etc [1].

CAIDA: Many ongoing projects, including Archipelago (ARK) for AS level topology updates. Also CAIDA Internet Data Catalog [2].

MoMe : EU FP6 initiative which gives a standardised interface to data from several ongoing projects and a data catalog [3].