Prototype Knowledge Base: an on-line information service in dependability and security

Hugh Glaser
Electronics & Computer Science
University of Southampton

Budapest, 22nd. March 2007

With

• Ian Millard
• Afraz Jaffri
• Benedicto Rodriguez
• ReSIST Partners
  – esp. Brian Randell
Background:
Semantic Web Challenge 2003 Winner

• CS AKTive Space
  – Gather data
  – UK People, projects, publications
    • Research funding
    • Top Universities
  – Geographical presentation

• AKT Project (www.aktors.org)

The Challenges

• Scientific Intelligence
  – Who is doing what where?
  – What impact are they having?

• Integrating resources
  – CORDIS, Institutional DBs and web sites, ePrints, NSF, CiteSeer, RIKS list, ISO LoCodes…

• Information: distributed and heterogeneous
  – Not under own control
  – Not in a common format
  – Not where you expect it

• Presenting to users & agents
ReSIST - Start Again

- A ReSIST Knowledge Base - The RKB
- Project Infrastructure support
- Europe (no longer UK-centric), the World
- Up to date
- Extra subject targets (resilience)
- Browser & platform independent
- Engineer for maintenance
  - Empower partners and other contributors
  - Empower other application builders

ReSIST - and deliver

- D10 - 2007-01-01T00:00:00A
- In fact it is just a URI to a service:
  - http://resist.ecs.soton.ac.uk/sparql/
- Or the raw content can be browsed
  - http://resist.ecs.soton.ac.uk/browse/
- But there is a brand new faceted browser
  - http://resist.ecs.soton.ac.uk/explorer/
- The RKB is embedded in the infrastructure
- The prototype is already being used
Structure

3store → CRS

3store

Conversion & Versionning

Ontologies, etc.,

Citeeer, CORDIS, DBLP, Partners, UN LoCode, …

Sources

- Publications
  - Partners
  - Citeeer
  - DBLP
  - ACM
  - DSN & FTCS Series
- Documents
  - RISKS Digest
- Projects
  - CORDIS
  - NSF
- People
  - Partners
- Support
  - UN LoCode
Ontologies etc.

- AKT Ontology
  - Scientific Research Activity
  - Dates
  - Location
  - ...
- ALRL Paper
- Courseware (extension of LOM)
- RISKS Codes
- ACM Classification

Main Browser - RKB Explorer

ReSIST RKB Explorer

Resilience for Survivability in IST

Network of Excellence in distributed Computing Systems
ResIST WG Algo
Network of Excellence in Distributed Comp
Basic Research on Advance
ResIST Training and Dissemination Coop
Assessment of QEN Technologies and their Application
Malicious and Accidental Fault Tolerance
European e-learning, dissemination project
ReSIST W3 Arch
Design for validation
ReSIST W3 Exec Board
ReSIST Resilience-Explicit Comp

Details

Name: Resilience for Survivability in IST
Funding source: The European Union
Funding amount: 4500000 EUR
Start date: 2006-01-01
End date: 2008-12-31
Course Metadata

ReSIST / Courses / Editing 'Advanced seminars on Distributed Systems'

- Name of the course: Advanced seminars on Distributed Systems
- Taught at:
  - Università degli studi di Bari, La Sapienza
  - Université catholique de Louvain
  - Université de Toulouse 1
  - University of Naples
  - Università degli studi di Napoli

- Currently being taught:

- Description:

  The course focuses on recent advances in distributed systems. A set of topics is selected and studied through the help of original papers. Practically, most known distributed system platforms are selected and analyzed.

- Language(s) of the course: English

Course Locations

Budapest University of Technology and Economics

Courses taught at Budapest University of Technology and Economics, Budapest:
- Software Verification and Validation
- Management of Computing Infrastructure

Go to the ReSIST Partner Map
Welcome to the ReSIST Wiki, which is the internal communication mechanism for the EU funded ReSIST "Network of Excellence" IP. Note that virtually all pages are private, and viewable only to ReSIST members who have logged in. Most content can be found by firstly browsing the main ReSIST page, which details the different research areas in which activities are ongoing as part of the project. If you have any questions or problems, please check that they have not previously been answered in the frequently asked questions, before contacting Ian Millard or Hugh Glaser at Southampton.

Quick Links
- Frequently asked questions
- ReSIST project page
- Recent changes to the wiki
- Upload new file / View uploaded files
- ReSIST members / photos / locations
- Calendar of Events
- Browse:0, query:0, or find out more about the Resilience Knowledge Base

Edit your User Interests

Please select the topics from within the hierarchy below that best match your research interests within the ReSIST NoE. It is best to "drill down" as far as possible, and to select the most specific topics. Selecting higher level topics will indicate that your are interested in all of the sub-topics, which are selected for you. Note however that this is not strictly a tree, as some topics appear in multiple places within the hierarchy. In these cases the "other" instances are automatically selected when you tick a topic area which exists in more than one category. As is usual within the wiki, clicking a blue link should take you to a page describing the subject of that link.

Happy clicking :)
Classifying

Manual classification of IEEE DSN papers

Authors: F. Jahanian, F. Jahanian, Z.M. Mao, E. Cooke
Abstract: Self-propagating malware like worms and bots can dramatically impact the availability and reliability of the Internet. Techniques for the detection and mitigation of Internet threats using content prevalence and scan detectors are based on assumptions of how threats propagate. Some of these assumptions have recently been called into question by observations of huge discrepancies in the quantity of specific threats detected at different points around the Internet. We call these deviations from uniform propagation “hotspots”. This paper quantifies and explains these influences on malware propagation. We then propose that hotspots can be explained by two fundamental influences on propagation: algorithmic factors and environmental factors. We use measurement data from sensors deployed at 11 locations around the Internet to demonstrate the impact of these factors on worm and bot propagation. With this understanding, we simulate the outbreak of new threats with hotspots and show how algorithmic and environmental factors reduce the visibility of distributed detectors resulting in the inability to identify new threats.

Keywords: None

Please select:
- adt:Research Area
  - Dependability And Security, Trustworthiness
    - Two notions overlapping concepts, with dependable long an integrating concept that encompasses the attributes: availability, reliability, safety, integrity and maintainability, while security encompasses confidentiality as well as integrity and availability.
  - Dependability, High Confidence, Survivability
    - The original definition of dependability is the ability to deliver service that can be trusted. The formal definition, that provides the criterion for deciding if the service is dependable, is the probability of a system is the ability to avoid service failures that are more frequent and more severe than acceptable.
  - Dependence
    - The dependence of system A on system B is the extent to which system A's dependability is or would be affected by that of system B.
  - Trust
    - Accepted dependence - when the dependence of a user on a given system

Browsing Raw Data

RKB Browser :: John Rushby

Identifiers...
http://catless.ncl.ac.uk/person#6060abae
http://citeeseer.ics.uci.edu/cgi-bin/search?w=John%20Rushby
http://citeeseer.ics.uci.edu/cgi-bin/search?w=John%20Rushby
http://citeeseer.ics.uci.edu/cgi-bin/search?w=John%20Rushby

Subject | Property | Object/Value
--- | --- | ---
John Rushby | akt:full-name | Rushby
JOHN RUSHBY | acm:proceedings | >>
John Rushby | acm:bookchapters | >>
John Rushby | acm:periodicals | >>
At the Centre
So what is RDF...?

- Resource Description Framework
- W3C recommendation
  - From Semantic Web research efforts
- Modelling language
  - Represents facts about resources
- Can model any abstract domain
  - Things do not have to be accessible on the web
  - But can be described in it

RDF: Basic components

- RDF graphs are formed by *triples*

```
http://laas.fr/people#laprie http://foo.com/example#email laprie@laas.fr
```
Important Components
3store and CRS

- 3store
  - Open source semantic store
  - Scalable
    - ReSIST - 50 million facts
      - (cf Wikipedia metadata)
- CRS - Consistent Reference Service
  - Bridges between disparate sources

Openness

- Almost nothing shown was private
- Except
  - Wiki project discussion pages
    - But semantic relations go to RKB
  - Data entry
    - Controlled
    - Not moderated
Future for ReSIST & the RKB

- Improve on the Prototype
  - Sources
  - CRS
  - UI
- Resilient-Explicit Computing
  - Model expert knowledge
  - Model processes, components, mechanisms
- Support Engineer/Scientist
  - Move effectively between
    - System design
    - Knowledge Base
    - People
  - To choose cost, characteristics, etc
- Support Run-Time Deployment
  - Dynamic Reconfiguration

Future Resources

- Original proposal
  - Now primarily maintenance
- Victim of success?
  - Important infrastructure
  - Serious resources to be maintained
  - People want to provide data (costs)
Response

- **ReSIST**
  - Has increased future RKB resources
- **Other Funding and Additionality**
  - Lithuania & Saarbrücken
  - JISC
- **Longer term**
  - Self-funding - SIGs, Clubs
  - Infrastructure - EU, EPSRC, NSF
- **Engineer for maintenance and Openess**
- **Open**
  - Knowledge Sources
  - Knowledge Publishing

Some Review Highlights

- One year of work - one RF funded
- ReSIST has done what it said it would do
  - And more
    - In particular, 1M -> 40M
    - Sophisticated UI
- Real tool for the network, from Day One
- Excellent Partner co-operation
  - Data
  - Evaluation
  - Ontology work
- Much Value in Expert Involvement