

SN15: Syxaw: XML-aware File Synchronization

Fuego Core Project / Helsinki Institute for Information Technology
Tancred Lindholm <tancred.lindholm@hiit.fi>

Vision

Syxaw uses links to build a web of synchronization relationships, just as hyperlinks have built the World Wide Web.

Central Ideas

- Syxaw synchronizes objects = files and directories
- Objects have associated *links*
- The link connects objects whose content is synchronized
- Data is always writable, even during network breaks
- Support for reconciliation of structured (XML) data

The Synchronization Link

- To share data we establish *synchronization links* between local files and directory trees on different devices
- Links are expressed as URIs
- The link is persistently stored along with other object metadata
- Links are *directed* and *local*: if *a* links to *b*, *b* does not know it.
- When synchronizing an object *a* linked to an object *b*, we reconcile changes made to these objects since the last point of synchronization
- In the case that the objects are directory trees, the full content, including contained objects, is synchronized.
- Links can contain *queries* to subset the synchronized data


Software Architecture

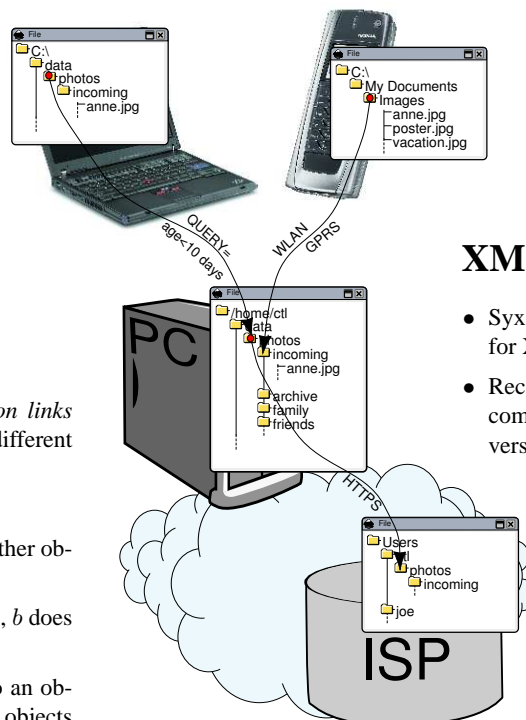
- Designed for limited devices, low bandwidth and high latencies
- Get interoperability for 10% penalty over most efficient solution
- XML used for data representation throughout
- Utilizes the Xebu optimized serialization
- Strive to be high quality software engineering

Implementation

- Implemented in 100% Java
- Open Source implementation at <http://hoslab.cs.helsinki.fi/savane/projects/fuego-core/>
- Ported to Java Foundation Profile in 2005
- Demoed on Linux PC, Nokia 9500 Communicator, Windows laptop.

The Figure

- Photo sharing scenario, with links drawn as 
- Photos are captured with a camera phone and synchronized over GPRS/WLAN to the home PC.
- Phone ↔ PC synchronization is set up by linking the Images folder on the phone to the incoming folder on the home PC.
- The laptop links to the PC to get photos less than 10 days old by attaching a *query* on the file age to the link
- The PC synchronizes with an ISP that provides secure archival storage



XML Reconciliation

- Syxaw entails a generic reconciliation framework for XML data
- Reconciliation by three-way-merging of previous common version, current local version and remote version
- Any conflicts are detected and kept local → conflicts do not deadlock the system
- To efficiently merge large XML documents, we replace XML subtrees with placeholder tags [1].
- Do as you teach: Directory trees are expressed as XML and reconciled using the XML reconciliation framework

Synchronization Protocol

- Implemented using HTTP POST and GET
- Efficient design which minimizes RPC roundtrips and batches RPC operations
- Only three kinds of RPC calls: get object, put object, and allocate object.
- Supports object deltas and compression
- Directory trees handled like any other objects on the protocol level → simple protocol
- Was found to be reasonably fast on a Nokia 9500 + GPRS

References

- [1] Tancred Lindholm, Jaakko Kangasharju, and Sasu Tarkoma. A hybrid approach to optimistic file system directory tree synchronization. In *Fourth International ACM Workshop on Data Engineering for Wireless and Mobile Access*, June 2005.