Peer-to-Peer Technology:
Where is it Heading

Prof. Dr. Techn. Mads Nygård

Department of Computer Science
Norwegian University of Science and Technology

Outline – Parts - Themes - Discussion

1. What is it – and What is it not?
2. Its Roots and History
3. Its Advantages and Disadvantages
4. Its Status of Today
5. What Aspects Might evolve?
6. Different Current Predictions
7. Own Personal Predictions
8. And for Collaborative Systems …

What is it – and What is it not?

- Definitions
  - “P2P is a class of applications that takes advantage of resources – storage, cycles, content, human presence – available at the edges of the Internet” (Shirky)
  - “A P2P system is a self-organizing system of equal, autonomous entities (peers) which aims for the shared usage of distributed resources in a networked environment avoiding central services” (Steinmetz)
  - “P2P is about overcoming the barriers to the formation of ad hoc communities, whether of people, of programs, of devices, or of distributed resources” (O’Reilly)

What is it – and What is it not? (cont.)

- Characteristics
  - Their design ensures that each user contributes resources to the system
  - Although they may differ in the resources that they contribute, all the nodes in a peer-to-peer system have the same functional capabilities and responsibilities
  - Their correct operation does not depend on the existence of any centrally administered systems
  - They can be designed to offer a limited degree of anonymity to the providers and users of resources
  - A key issue for their efficient operation is the choice of an algorithm for the placement of data across many hosts and subsequent access to it in a manner that balances the workload and ensures availability without adding undue overheads

What is it – and What is it not? (cont.)

- Left out – Without or With Intention !?
  - The dynamic set of nodes in the system
  - The arbitrary scaling ability
  - The server-less situation
  - The symmetric situation
  - The overlay routing
  - The centralized remains

Its Roots and History

- Generations - Searching
  1. Gen.: Centralized servers (…)
     Napster
  2. Gen.: Request flooding (Unstructured)
     Freenet, Gnutella, Kazaa, BitTorrent
  3. Gen.: Peer-to-peer Middleware (Structured)
     Pastry, Tapestry, CAN, Chord, Kademlia
### Its Advantages and Disadvantages

- **Advantage (Today)**
  - Extreme Scalability!
  - Suitable Reliability!
  - Decent Fault-tolerance!

- **Disadvantage (Today)**
  - Too much Freedom?
  - How much Security?
  - To little Anonymity?

### Its Status of Today

- **Partly Cycle Sharing?**
- **Mostly File Sharing!**
- Audio & Video …
- Manually Tagged …
- Popularity!
- Anonymity?

### Its Status of Today (cont.)

- **Peer-to-Peer Middleware – Func. Requirements**
  (Coulouris, Dollimore & Kindberg 05)
  - Simplify the construction of services implemented across many hosts in a widely distributed network
  - Enable clients to locate and communicate with any individual resource made available to a service
  - Accept adding and removing hosts to and from the system
  - Enable adding and removing resources to and from a service at will
  - Offer a simple programming interface

### Its Status of Today (cont.)

- **Peer-to-Peer Middleware – Non-func. Requirements**
  (Coulouris, Dollimore & Kindberg 05)
  - Global scalability
  - Load balancing
  - Optimization for local interactions between neighbouring peers
  - Accommodating to highly dynamic host availability
  - Security of data in an environment with heterogeneous trust
  - Anonymity, deniability and resistance to censorship

### Its Status of Today (cont.)

- **Overlay Routing**
  - Distributed Hash Table (DHT)
    - Pastry
    - Squirrel
  - Distributed Object Location and Routing (DOLR)
    - Tapestry
    - OceanStore
  - Others …

### What Aspects Might evolve?

- **Dimensions**
  - Unstructured - Structured
  - Layering - Caching
  - Pulling - Pushing
  - Professional - Personal
  - ReadOnly - WriteAlso
  - Partitioning - Replication
  - Availability - Security
  - Anonymity - Replication
  - Encryption - Control
Different Current Predictions (new)

- Rabitti 06
  - Requirements of Future Internet-based Applications:
    Scalability, security, flexibility
  - In the Future:
    Explosion of Data Creation - Mostly in Audio/Visual Form
  - New Searching Paradigm:
    Similarity Search
  - Non-P2P Systems:
    Non-scalability requires Approximate Techniques
  - P2P Systems:
    Similarity Search remains Scalable

Different Current Predictions (old)

- Crowcroft & Pratt 02
  - Measuring and Modelling P2P Systems
  - Update and Persistence in P2P Storage Systems
  - Computation
  - QoS and Accounting
  - Locality versus Anonymity
  - Evolution of P2P into Lower Layers
  - P2P and Ad-hoc Wireless Network Duality

Different Current Predictions (...)

- Coulouris, Dollimore & Kindberg 05
  - Calls for better solutions:
    That the storage they use for mutable data is relatively costly compared to a trusted, centralized service
  - Calls for better approaches:
    That the promising basis they provide for client and host anonymity has not resulted in strong guarantees of anonymity

Different Current Predictions (...)

- Tanenbaum & van Steen 07
  - Unstructured - Structured
  - Caching - Replication
  - Availability
  - Security

- Rabitti 06 (cont.)
  - Possible & Promising:
    - Advanced Distributed Index Structures
    - Cooperative Audio/Visual Features Extraction
    - Push-based Crawling – Providers to Indexing Nodes
    - Cooperative Crawling – Hidden/Controlled Multi Media Content
    - Support of Highly Dynamic Applications
  - Research challenges:
    - Complex Similarity search with Different measures and Multiple overlays
    - Dynamic Load balancing including Search collaboration
    - Dynamic Caching and Replication of A/V contents, Query results and Index entries
    - Digital Rights Management from Audio watermarking to Song similarity

- Crowcroft & Pratt 02 (cont.)
  - Study of P2P Systems:
    - Partly sociological
    - Partly ecological
    - Partly economical
    Small World Model

- Different Current Predictions (...)

  - Indirectly Given
Own Personal Predictions

- Research needs / Development trends
  - Anonymity/Encryption: Because it concerns Basic Needs
  - Digital Rights Management: Because it concerns Monetary Aspects
  - Overlay Routing: Because it concerns Effectiveness
  - Caching and Replication: Because it concerns Effectiveness
  - Data Management: Because it concerns P2P’s Eventual Maturity
  - Distributed Computing: Because it concerns P2P’s Inherent Option

And for Collaborative Systems …

- Most potential:
  - Collaborative Crawling Model (Rabitti 06)
  - Collaborative Business Model (Rabitti 06)
  - Collaborative Framework Issues (Le & Nygård 05)
  - Collaborative Transactional Aspects (Nygård & Le 06)

References (1:2)


References (2:2)