



## Supporting Information Retrieval in Peer-to-Peer Systems

Wolf-Tilo Balke, Klara Nahrstedt, Heiko Niedermayer

Schloß Dagstuhl, 28.03.06



## IR support for Peer-to-Peers infrastructures

- Multiple kind of information
  - Data/files with simple metadata
  - Complex text documents
  - Annotated data, process/system state information, log files,...

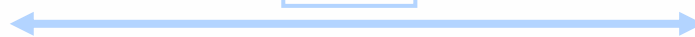
database  
techniques

information retrieval  
techniques

data/files

log files

complex  
documents





## Application domains

- Long living multimedia applications:
  - long connection times (video streaming ~2 hours+), QoS information, multiple copies (MPEG-21), frame numbers, verbose metadata descriptions,...
- Security applications:
  - logs of firewalls, connection logs (intrusion detection), number of flows, bandwidth patterns, header fields, priorities, trust metadata,...
- Large scale distributed experiments:
  - Intermediate results, states, counters, error codes and messages, service descriptions, service functionality, service parameters,...



## Requirements on IR and P2P

- Large and heavily distributed data items
- Support for fuzzy queries, range queries, and ranked query models
- Queries on aggregated information, correlations, conditional expressions over system parameters,...
- Graceful recovery from application failures, system failures, network churn, ...
- Reliability of derived information: precision, recall,...
- Online-algorithms and real-time IR techniques

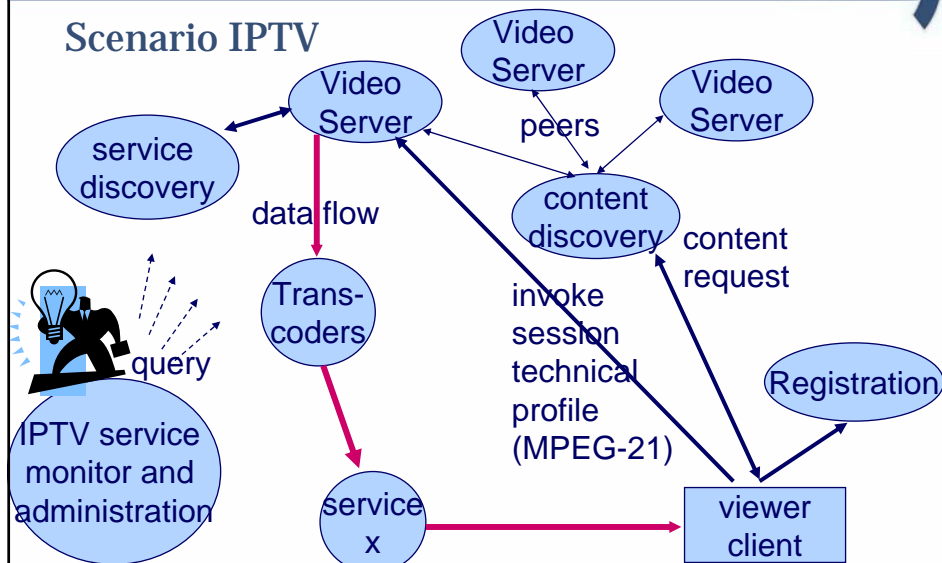


## Challenges

- Data stream vs. control plane
- Organization and indexing of files
  - Types of indexes
  - Indexing highly discriminative information (only/preferred)
  - Resolution of indexing (ontologies, taxonomies,...)
  - Compressing/aggregating past vs. recent vs. current information to save space
  - Creation of dictionaries, controlled vocabularies,...
  - Clustering of characteristics, memberships in sets
- IR operations
  - Search, filtering, latent semantics, recommendation, reputation, membership functions for overlay construction/community building, precision/recall vs. efficiency, caching/memoization



## Scenario IPTV





## 2. International Workshop on Multimedia Service Composition

- <http://www.l3s.de/msc06>
- October, 28<sup>th</sup> in Santa Barbara, CA
- All problems considered
  - Research papers (up to 10 pages)
  - Application/experience papers (up to 10 pages)
- Deadline early July



**Questions?**

**balke@l3s.de**

