Philippe Cudré-Mauroux
and Karl Aberer

Message Passing in Semantic Peer-to-Peer Overlay Networks

Peer-to-peer (P2P) systems rely on machine-to-machine ad-hoc communications to offer services to a community. Contrary to the classical client-server architecture, P2P systems consider all peers, i.e., all nodes participating in the network, as being equal. Hence, peers can at the same time act as clients consuming resources from the system, and as servers providing resources to the community. P2P applications function on top of existing routing infrastructures, typically on top of the IP network, and organize peers into logical and decentralized structures called *overlay* networks.

In this column, we discuss exploratory research related to data management in P2P overlay networks. First, we discuss the notions of unstructured and structured P2P overlay networks. Then, we discuss data management in such networks by introducing an additional layer to handle semantic heterogeneity and data integration. Finally, we present a method based on sum-product message passing to detect inconsistent information in this setting.

P2P OVERLAY NETWORKS—ARCHITECTURE

Increasingly on the Internet, applications are supported by sets of loosely connected machines operating without any form of central coordination; Internet telephony networks such as Skype [1] and file sharing applications like Gnutella [2] are two well-known examples of this trend. Contrary to the client-server setting, where applications are bound to sets of static servers identified by an IP address, these applications need ways of organizing the dynamic sets of machines providing the service. P2P overlay networks address this need and allow the management of virtual and decentralized networks created on top of the IP infrastructure.

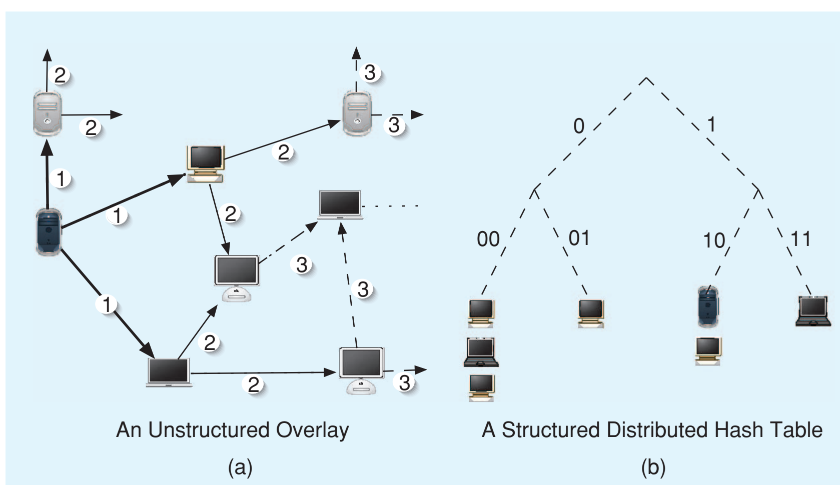
The virtual structure connecting all the peers operating in an overlay network can vary. In *unstructured* overlay networks such as Gnutella, peers establish connections to a fixed number of other peers, creating a random graph of P2P connections. Requests originating from one peer are forwarded by the other peers in a cooperative manner, as depicted in Figure 1(a). This relatively simple and robust mechanism is, however, network-intensive, as it broadcasts all queries to all peers within a certain radius irrespective of the content of the query.

Structured overlay networks were introduced to alleviate network traffic while maximizing the probability of a query locating a specific peer. Peers in a structured overlay can for example be organized on a multidimensional torus [3] or into a virtual binary search tree, as promulgated by the P-Grid P2P system [4] and illustrated in Figure 1(b). Such systems provide hash-table functionalities on an Internet-like scale and are known as distributed hash tables (DHTs). They typically enable global search on shared data items in a totally decentralized way in $\mathcal{O}(\log(N))$ messages (i.e., packets sent from one peer to another), where N is the number of peers in the overlay.

P2P OVERLAY NETWORKS—DATA MANAGEMENT

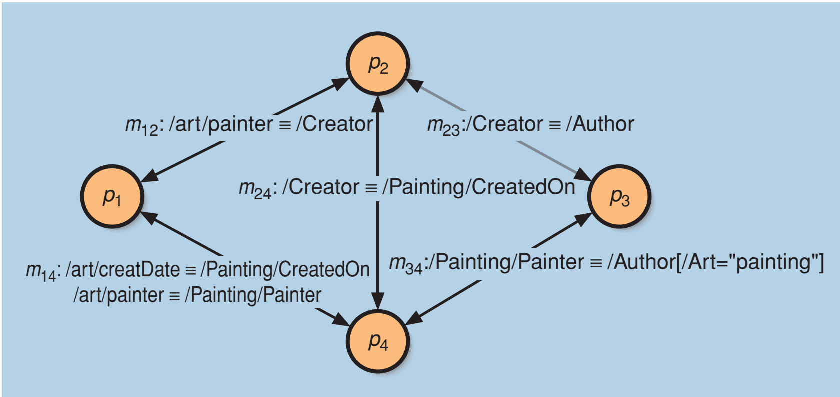
SEMANTIC HETEROGENEITY

P2P overlays originally dealt with very simple data and query models: only file names were shared and queries were composed of a single hash value or a keyword. Rapidly, several research efforts [4] tried to enrich overlay networks with more expressive models to support structured data conforming to schemas. In



[FIG1] **Two P2P architectures: (a) an unstructured P2P overlay a la Gnutella: a query originating from a peer on the left-hand side of the figure is iteratively gossiped up to three times, and (b) a structured distributed hash-table a la P-grid: peers are organized into a virtual binary tree.**

Digital Object Identifier 10.1109/MSP.2007.904799

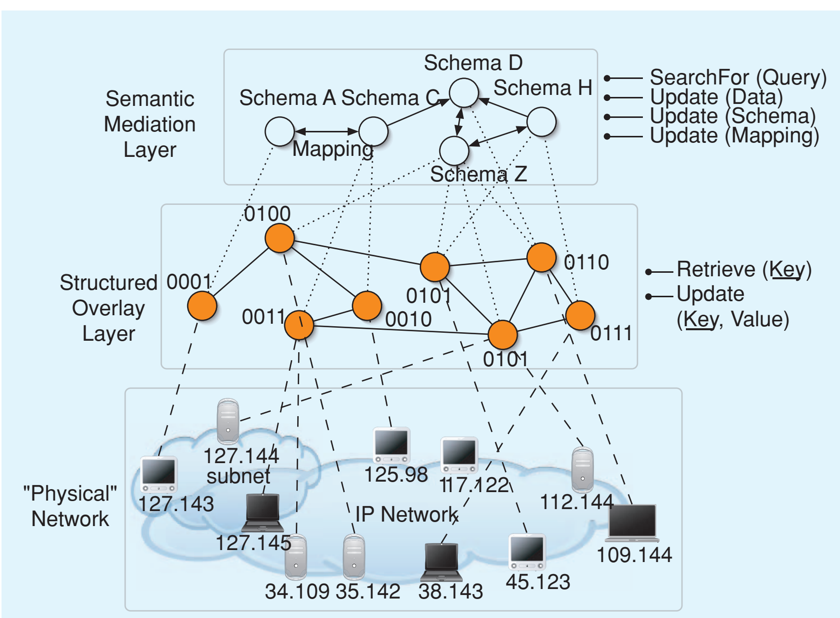


[FIG2] A simple example of a peer data management system; four peer databases *p* are connected through five pairwise schema mappings *m*, used to reformulate queries originating from one database into corresponding queries at the other databases.

data management, schemas are used as declarative models to define the structure of the data. Relational schemas, for example, organize data in n-ary relations (i.e., tables), while XML schemas structure data through hierarchies of elements.

Sharing (semi) structured data, such as relational tuples or XML instances, is however intricate in decentralized and P2P networks. Rapidly, the problem of semantic heterogeneity surfaces, as distinct sources might encode similar information using different schemas to structure their data.

To tackle the problem of semantic heterogeneity and relate similar pieces of information encoded using different schemas, database systems take advantage of schema *mappings*. Mappings provide a bridge between the definitions of two schemas. They allow to reformulate a structured query, such as an SQL query, posed against one schema into a new query posed against a similar schema. A centralized component called a *mediator* [5] would typically store a global schema that is related to all the schemas of the databases through mappings. Thus, the mediator provides a central query interface that allows transparent access to disparate and heterogeneous database systems.



[FIG3] A semantic P2P overlay network: in many practical settings, the semantic mediation layer relating the schemas is independent from the organization of the peers, which is itself dissociated from the physical network structure of the machines.

PEER DATA MANAGEMENT

Maintaining a global schema integrating all data sources is not desirable in decentralized infrastructures such as P2P systems where peers are loosely coupled. Recently, a new breed of database systems called *peer data management systems* (PDMSs) [6] emerged as an attempt to decentralize the mediator architecture and allow the systems to scale gracefully with the number of heterogeneous sources.

Instead of requiring the definition of a global schema, PDMSs consider loosely structured networks of mappings between pairs of schemas to iteratively disseminate a query from one database to all the other related databases. A simple example of a PDMS network is illustrated in Figure 2, where four peer databases $p_1 \ldots p_4$ are connected through five schema mappings. The mappings, which are usually expressed as queries, can be used to reformulate a query iteratively to disseminate the query throughout the network. Once a database sends a query to its immediate neighbors through local mapping links, its neighbors (after processing the query) in turn propagate the query to their own neighbors, and so on and so forth until the query reaches all (or a predefined number of) databases. The way the query spreads around the network mimics the way messages are routed in an unstructured P2P system such as that shown in Figure 1(a).

For a new peer database wishing to join an existing PDMS, the cost of entry is minimal as for most P2P systems: the new peer only has to define a few schema mappings between its schema and the schemas of other peers already connected to the system. The peers can continue to handle their data using their own schemas, but can query all the other databases thanks to iterative query reformulation through the mappings. In case of an intermediate node failure, peers can reroute their queries through different schema mappings or create new mappings to circumvent the offline peer and continue to query distant sources.

SEMANTIC P2P OVERLAY NETWORKS

In practice, a single schema can be used simultaneously by many independent parties. Furthermore, some peers might choose more than one schema to structure their data locally, as they realistically might have to handle very diverse pieces of information. Hence, the organization of the schemas and mappings can often be uncorrelated with the organization of the peers themselves. A model in which physical machines form a P2P overlay network, which is itself independent of the semantic schema-to-schema overlay handling data integration, is illustrated in Figure 3.

In this figure, the base layer represents an IP network with different machines identified by IP addresses. The machines self-organize into a virtual P2P overlay network (middle layer), with its own addressing scheme. The overlay, which is typically structured, only supports very simple operations (*Retrieve*, *Update*) on hash values. The upper layer consists of the structured data, schemas and semantic mappings relating the schemas. It is created using the overlay layer to collaboratively share all semantic information, allowing all peers at the overlay layer to query and update the semantic mediation layer, and reformulate queries following different semantic paths.

INCORRECT SEMANTIC MAPPINGS

In semantic P2P overlay networks, consistency or quality of the schema mappings stored in the semantic mediation layer is one of the main concerns. Incorrect mappings generate incorrect reformulations of the queries, and thus incorrect results. As P2P systems target large-scale, decentralized, and heterogeneous environments where autonomous parties have full control on the design of their own schemas, it is not always possible to create correct mappings between two given schemas. As a result, the creation of correct mappings is often precluded. Therefore, in many cases an approximate mapping relating two similar but semantically slightly divergent concepts might be more beneficial than no mapping at all.

Many different techniques, based on classification or data mining, can be used to map schemas automatically and to evaluate the correctness of the mappings [7]. In what follows, we focus on the semantic mediation layer and propose a radically new technique that takes advantage of the distributed nature of the schemas and mappings in the system to detect the mappings whose semantics diverge from the semantics of other mappings.

ANALYSIS OF RETURNING QUERIES

When reformulating a query using series of mappings, cycles or redundant paths of reformulations can occur (for example going from p_1 to p_2 to p_4 and back to p_1 in the simple semantic mediation layer depicted in Figure 2). When such a cycle is detected, a peer can compare the original query q it received to the reformulated query q'. Two cases can occur:

$q' \equiv q$: this occurs when the query, after having been reformulated n times through n mappings, still is equivalent to the original query q. Since this indicates a high level of semantic agreement along the cycle, we say that this represents positive feedback f^+ on the mappings constituting the cycle.

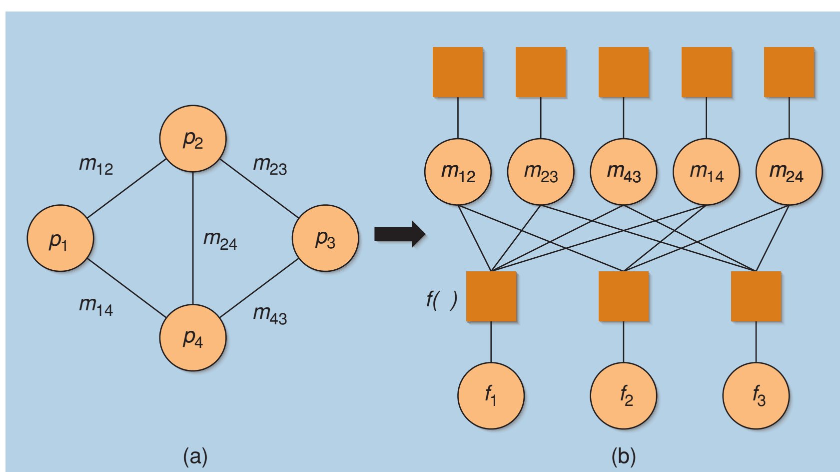
$q' \neq q$: this occurs when the query, after having been reformulated n times through n mappings, returns to a peer as a semantically different query. As this indicates some disagreement on the semantics of the query along the cycle of mappings, we say that this represents negative feedback f^- on the mappings constituting the cycle.

ANALYSIS OF THE SEMANTIC MEDIATION LAYER

Series of mappings $m_1, m_2 \ldots m_n$ can accidentally compensate their respective errors and actually create a semantically correct composite mapping $m_n \circ m_{n-1} \ldots \circ m_1$ in the end. Assuming a probability Δ of two or more mapping errors being compensated along a cycle we can determine the conditional probability of a cycle producing positive feedback $f^+_\circlearrowleft$ given the correctness of its constituting mappings $m_1, \ldots, m_n$:

$$P(f^+_\circlearrowleft | m_1, \ldots, m_n) = \begin{cases} 1 & \text{if all mappings correct} \\ 0 & \text{if one mapping incorrect} \\ \Delta & \text{if two or more} \\ & \text{mappings incorrect.} \end{cases}$$

Examining again Figure 2 posing a query $q = /art/painter$ at p_1 and sending the query to p_2, p_4 and then back to p_1 would result in a returning query $q' = /art/creatDate$. Obviously, the semantics of the query have changed after having been reformulated three times. The question is which mapping is responsible for that change. It is difficult



[FIG4] Modeling (a) an undirected network of mappings as (b) a factor graph ; the nodes of the factor graph represent, from top to bottom, probability functions encapsulating *a priori* information on the schema mappings, variables for the semantic correctness of schema mappings, probability functions relating schema mapping variables to feedback variables, and finally variables for the feedback that can be observed in the network.

to give an answer for a single cycle. Instead, one would need to consider several cycles and their correlation through their shared mappings.

Feedback can be obtained from three different mapping cycles in the network depicted in Figure 2:

$$f^1_{\circlearrowleft}: m_{12} - m_{23} - m_{34} - m_{41}$$
$$f^2_{\circlearrowleft}: m_{12} - m_{24} - m_{41}$$
$$f^3_{\circlearrowleft}: m_{23} - m_{34} - m_{24}.$$

One might construct a global probabilistic graphical model linking the mappings and the feedback information from the network. Figure 4 depicts a factor graph [8] for the small network of Figure 2, containing (from top to bottom): five one-variable factors for the prior probability functions on the mappings, five mappings variables m_{ij}, three factors $f()$ linking feedback variables to mapping variables through conditional probability functions (defined earlier), and finally three feedback variables f_k. Note that feedback variables are usually not independent (e.g., in Figure 4, all three feedbacks are correlated).

DISCOVERY OF ERRONEOUS MAPPINGS

Posterior values on the semantic correctness of the mappings given the feedback observed in the network can be computed using standard sum-product message passing techniques [8]. Such computations can be performed globally, for example by deriving a junction tree from the factor graph, or can be approximated locally by the peers using iterative message passing schemes. Given a local factor graph containing only the mappings it is responsible for, a peer can locally update the believes on its mappings by reformulating the sum-product algorithm. Messages M from mapping variables m to factors $f()$ can be sent repeatedly, both locally to the local factor $f()$, and remotely to peers responsible for the other mappings connected to $f()$. Using $f()$ as a function returning the neighboring nodes of a given node in the factor graph and $\sum_{\sim\{x_i\}}$ to denote a summary over all but one variable x_i:

local message from local mapping m_i to factor $f_j() \in n(m_i)$:

$$M_{m_i \to f_j()}(m_i) = \prod_{f() \in n(m_i) \setminus \{f_j()\}} \times M_{f() \to m_i}(m_i)$$

remote message from local mapping m_i to factor $f_k()$ from peer p_0 to peer p_j responsible for a mapping $m \in n(f_k())$:

$$M_{p_0 \to f_k()}(m_i) = \prod_{f() \in n(m_i) \setminus \{f_k()\}} \times M_{f() \to m_i}(m_i).$$

Messages for the mapping variables can then be computed by combining both local messages and remote messages received from distant peers:

local message from factor $f_j()$ to mapping variable m_i:

$$M_{f_j() \to m_i}(m_i) = \sum_{\sim\{m_i\}} \left(f_j() \prod_{p_k \in n(f_j())} M_{p_k \to f_j()}(p_k) \times \prod_{m_l \in n(f_j()) \setminus \{m_i\}} M_{m_l \to f_j()}(m_l) \right)$$

posterior semantic correctness of local mapping m_i:

$$P(m_i | f) = \alpha \left(\prod_{f() \in n(m_i)} M_{f() \to m_i}(m_i) \right)$$

where α is a normalizing constant ensuring that the probabilities of all events sum to one (i.e., making sure that $P(m_i = \text{correct}) + P(m_i = \text{incorrect}) = 1$). Examining again the simple semantic in Figure 2 and taking into account the feedback given by the three cycles appearing in the network, this analysis converges after a few iterations. Without any a priori information on the mappings, it successfully detects the inconsistencies related to the mapping between p_2 and p_4 ($P(m_{24} = \text{correct}) = 0.3$) while assessing the other mappings as correct ($P\text{correct}) = 0.6$).

Note that the allocation of responsibilities between the peers at the overlay layer and the semantic information at the mediation layer might be quite different in the general case of a three-layered system such as the one depicted in Figure 3. The calculations pertaining to some variables or factors might be redundant (as above, where each feedback variable is taken into consideration by several peers). Those computations can be embedded in the normal behavior of the network by piggybacking on the query reformulation traffic [9]. Several large-scale experiments were conducted in order to determine the performance of this technique [10]. The results are very promising for a totally automated technique, as the precision of the decision process in detecting correct or incorrect mappings ranges 80–100% for most realistic settings.

CONCLUSIONS

We presented a method that exploits a probabilistic model to assess the degree of semantic correctness of schema mappings in a semantic P2P overlay network. The method takes advantage of transitive closures of mapping operations to compare reformulated queries at the semantic mediation layer. Probabilistic values on the correctness of the mappings are then computed by iteratively passing messages between the peers.

Information integration techniques were traditionally centered around global schemas, perfect schema mappings, and static query rewritings. Those techniques are today inappropriate to maximize the performance of large-scale information systems that operate without any form of central coordination, such as semantic P2P overlay networks. Data noise—emerging from uncertainty on distant data or from a lack of coordination among the sources—plays an ever increasing role in structured data management. Correctly handling that new type of noise is an important challenge, that might well be tackled by techniques relating to signal processing and information theory, as we tried to outline with the message passing scheme presented above.

ACKNOWLEDGMENTS

This work was supported by the Swiss NSF National Competence Center in Research on Mobile Information and Communication Systems (grant 5005-67322) and by the EPFL Center for Global Computing as part of the European project NEPOMUK No FP6-027705.

AUTHORS

Philippe Cudré-Mauroux (philippe.cudre-mauroux@epfl.ch) is a senior researcher at Ecole Polytechnique Fédérale de Lausanne (EPFL) in Switzerland. His research interests are in information management in large-scale settings with an emphasis on semantic overlay networks and peer data management. He is a member of the IFIP Working Group 2.6 on Databases.

Karl Aberer (karl.aberer@epfl.ch) is a professor at EPFL and the director of the Swiss National Research Center for Mobile Information and Communication Systems. His research interests consist of the foundation, algorithms and infrastructures for distributed information management including peer-to-peer overlay networks, semantic interoperability, trust management and applications to scientific and sensor data management.

REFERENCES

[1] S. Guha, N. Daswani, and R. Jain, "An experimental study of the Skype peer-to-peer VoIP system." Available: http://saikat.guha.cc/pub/iptps06-skype/

[2] M. Ripeanu, "Peer-to-peer architecture case study: Gnutella network," in *Proc. Int. Conf. Peer-to-Peer Comput.*, 2001.

[3] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, "A scalable content-addressable network," in *Proc. ACM SIGCOMM*, 2001.

[4] K. Aberer (ed.), "Special issue on peer to peer data management," in *Proc. ACM SIGMOD Record*, vol. 32, no. 3, 2003.

[5] G. Wiederhold, "Mediators in the architecture of future information systems," *IEEE Computer*, vol. 25, no. 3, 1992.

[6] A.Y. Halevy, Z.G. Ives, D. Suciu, and I. Tatarinov, "Schema mediation in peer data management systems," in *Proc. Int. Conf. Data Eng.*, 2003.

[7] E. Rahm and P.A. Bernstein, "A survey of approaches to automatic schema matching," *Int. J. Very Large Data Bases*, vol. 10, no. 4, 2001.

[8] F.R. Kschischang, B.J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, 2001.

[9] P. Cudré-Mauroux, K. Aberer, and A. Feher, "Probabilistic message passing in peer data management systems," in *Proc. Int. Conf. Data Eng.*, 2006.

[10] P. Cudré-Mauroux, "Emergent Semantics: Rethinking Interoperability for Large-Scale Decentralized Information Systems," Ph.D. dissertation, 3690, EPFL, Lausanne, Switzerland, 2006. [SP]

[reader's **CHOICE**] continued from page 10

TITLE, AUTHOR, PUBLICATION YEAR	ABSTRACT	RANK IN IEEE TOP 100 (SEPTEMBER, 2006–FEBRUARY 2007) FEB	JAN	DEC	NOV	OCT	SEP	*N* TIMES IN TOP 100 SINCE JAN 2006
IEEE SPS JOURNALS								
LOSSLESS WATERMARKING FOR IMAGE AUTHENTICATION Celik, M.U.; Sharma, G.; Tekalp, A.M. *IEEE Transactions on Image Processing*, vol. 15, no. 4, Apr. 2006, pp. 1042–1049	This article presents a novel framework for lossless watermarking, which enables zero-distortion reconstruction of the un-watermarked images upon verification.						97	2
IEEE SPS CONFERENCES								
FULLY UTILIZED AND REUSABLE ARCHITECTURE FOR FRACTIONAL MOTION ESTIMATION OF H.264/AVC Chen, T.C.; Huang, Y.W.; Chen, L.G. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, May 2004, pp. 9–12	This conference article describes a new VLSI architecture for fractional motion estimation in the H.264/AVC video compression standard, which can also be used with other standards.					35		1

[SP]