Advertising and discovering trusted resources within a P2P-based architecture

Vincenza Carchiolo, Alessandro Longheu, Michele Malgeri and Giuseppe Mangioni

Dipart. di Ingegneria Informatica e delle Telecom. - Universita’ di Catania - V.Ie A. Doria, 6 - Catania (Italy)

E-mail {car, alongheu, mm, gmangioni} @ dii.unict.it

Abstract—The use of P2P approach in E-Learning is an interesting solution in particular within lifelong-learning. Generally, peer-to-peer paradigm is a successful solution to the problem of resources sharing. In this field there are two open questions. First, how to advertise Learning Objects on the network, so that peers knows they exist. Second, how to choose the best set of learning objects. The system presented in this paper aims at answering these questions.

I. INTRODUCTION

Life–long learning is the last frontier of e–learning, since the rapid evolution of technologies imposes to gather information from anywhere to understand today’s world. In this context, two aspects assume a central role. First, in the context of life–long learning the course personalization is essential to provide each student with a tailored course. Second, the decentralization is becoming more and more dominant thus promoting the use of Peer-to-Peer (P2P) networks for Learning Objects (LO) sharing. Such networks impose to address the question of how to trust distributed LOs in order to know whether they (and their providers) can be considered reliable or not; this is becoming more and more critical since students no more access a centralized system with materials managed by a single and (possibly) trusted authority, rather today the central role is played by the so–called Web 2.0 with the underlying infrastructure often based on the P2P paradigm [1].

The reliability of both LOs and their provider is a question we try to answer by exploiting the concept of trust [2]. A simple and effective definition for trust is “trust provides information about with whom we should share information, from whom we should accept information, and what consideration such information should be given” [3]. Another problem arises when the number and the dissemination of LOs both increases, making it harder to find the most adequate materials.

We then propose a decentralized architecture based on P2P network in order to improve LOs sharing and to this purpose we focus on trust management of LOs and on an advertising strategy used to disseminate the knowledge about which LOs are stored by which peers across the network. Moreover, we exploit PROSA [4] overlay unstructured network architecture inspired by social relationships, which provide us with significant features, indeed the simple reproduction of mechanisms observed in real world social networks allows peers to self organize, creating a network with small–world characteristics, and with an efficient and fast answer queries with high recall [5]. We test our proposal on an e–learning network built exploiting real data sets.

Current e–learning systems focus on both the improvement of LOs sharing and reusing and on the creation of tailored learning paths, but neither trust nor advertising seems to have been fully investigated.

An interesting project is endorsed by the Ariadne [6] foundation, which proposes the Knowledge Pool System to manage and reuse learning resources. ALFanet [7] is an adaptive e–learning system providing support in users adaptation for both information and contents format. ALFanet allows authors to add several contents annotating them with characteristics user should possess to access them (prerequisites). Each course is supported by a tutor that helps users and - whenever needed - adapts contents and formats. Presence of tutor cannot be though considered a general solution because several scenarios cannot effectively support them - i.e. life–long learning.

Edutella [8], a p2p network designed for distributed environments. The project defines a framework that provides mechanism to manage peer access, a reference “Edutella Common Data Model” allowing the sharing of queries and messages, but each peer refers only to local data representation and query model. Finally, we note that other issues have to be considered in P2P distributed scenario, e.g. the interoperability among different LOs repositories. [9] proposes an universal interoperability layer (named SQI) for educational networks, whereas in [10] reports an experience about Edutella and Ariadne interaction; however, in our work we do not focus on such questions. The paper is organized as follows: first we exploit PROSA architecture into the e–learning context, adding trust and personalization, then we present advertising mechanism. Finally, we show preliminary results of some simulations performed over a real data set.

II. ARCHITECTURE

To provide e–learners with as many trusted learning paths as possible to choose from, we introduce a P2P based architecture of prosumers (peers) and resources (LOs), where trust, personalization and advertising play a central role.

In particular, we exploit trustworthiness for both peers–LOs as well as for peer-to-peers links; peers are connected themselves into a trust network, so trust relationships among peers allows to select which ones of them can be considered more authoritative in answering a query within a given topic (described by shared ontologies). Resources are pointed by peers, i.e. providers or users that accessed that resource, to express the quality (trust) assigned to that resource, helping in
choosing the best ones during the learning process. Moreover, the personalization of learning paths is achieved by exploiting proper information stored into resources (e.g., the required knowledge) and into peers-resources links (e.g., the time required to learn the pointed resources as estimated by each pointing peer); such information will be shown later. Finally, the advertising mechanism allows to publicize the existence of resources over all the network, enhancing the set of reliable (trusted) learning paths to be exploited.

A. Using PROSA network in e-learning context

Starting to model the P2P network, we think it is natural to get inspiration from social networks since prosumers that establish their mutual trust relationships and assign trust to resources are actually humans beings; hence, we exploit PROSA, a P2P overlay unstructured network architecture inspired by social relationships described in [4]. In order to adapt PROSA to the e-learning context, we must provide additional information to build an effective and efficient e-learning system with trust and personalization. To this purpose, first we detail the generic PROSA resource as $R = \{ n_R, d_R, pre_R, obj_R \}$ where $n_R$ is a name, $d_R$ a description and $pre_R$ and $obj_R$ the set of prerequisite and objectives, that are those concepts respectively needed before and gained after learning that resource; they are both needed to personalize the sequence of resources to be learned according to the knowledge owned (prerequisites) and required (objectives) by the requesting peer. Considering links, those between peers are the same modeled by PROSA; in addition to general PROSA links, each link label must include the trustworthiness assigned by source to the target node; as frequently adopted in literature, trust can be modeled as a number in the range $[0, 1]$. To complete the definition for other links (i.e., peer-to-resource and resource-resource), we define the following:

- $T = \{ (A, (l^A_Y, time_Y, diff_Y), Y) | (A \in \mathcal{P}), (Y \in \mathcal{R}), (l^A_Y \in [0, 1]), (time_Y > 0), (diff_Y \in \{low, mid, high\}) \}$, as the set of peer-to-resource links, with $l^A_Y$ being the trust $A$ assigned to the resource $Y$. The $time_Y$ represents the estimated time needed to learn that resource’s contents, while $diff_Y$ represents the difficulty in learning resource’s contents, expressed as a value belonging to the set $\{low, mid, high\}$; trust, time and difficulty are all used to provide personalization.
- $S = \{ s_{A_X} : (A, X) | (A \in \mathcal{R}) \land (X \in \mathcal{R}) \land (A \neq X) \}$, as the set of precedence-succession relationships among resources, which express the learning path to be followed through a set of resources, i.e. the edge $s_{A_X}$ from $A$ to $X$ means that the resource $A$ should be learned before the resource $X$.

We use these information when running the PROSA management algorithm (see [4] for details), in particular:

- in the generic PROSA query message $Q_M$, the query $q$ issued by a learner peer $L$ is specified as $q = \{ pre_L, obj_L, time_L, diff_L, trust_L \}$ where $pre_L$ and $obj_L$ are prerequisites and objectives of the requesting peer, $time_L$ represents the desired total time available for the learner and $diff_L$ is the mean level of difficulty required; such needs will provide learner’s personalization in the learning path. Finally, $trust_L$ is the threshold for trustworthiness to be used during the search algorithm.
- in PROSA, the function $ResRelevance$ evaluates the relevance of a query with respect to the resources hosted by a peer. This function here is implemented by exploiting all values inside query $q$ plus values on edge sets $T$ and $S$ introduced above, in order to evaluate the relevance of hosted resources according to precedence-succession relationships together with time, difficulty and trust personalization constraints.
- the function $SelectNextPeer$ exploits trust as the weight used to evaluate the relevance of query $q$ for each neighbour; similarly, trust is used in the semantic flooding to select relevant neighbours the query should be forwarded to.

B. Exploiting advertising resources mechanism

As most unstructured P2P networks, in PROSA a new resource shared by a peer is not known to the other peers until it is requested and, eventually, downloaded. So, searching a recently shared resource could be very hard if there isn’t a priori known about resource location. Hence an advertising method for PROSA is presented, aiming at making known new shared resources to a given (carefully chosen) subset of PROSA peers. The method takes advantages of the underlying PROSA query management algorithm, exploiting knowledge representation mechanism to represent an advertise. In particular, an advertisement message is represented by a tuple $A_M = (aid, a, s_{A}, ttl)$, where $aid$ is a unique identifier used to avoid multiple processing of the message by the same peer, $a \in Q$ is an advertisement defined in the Query Space, $s_{A} \in \mathcal{P}$ is the source peer (i.e. the peer owning publicized resources), $ttl \in \mathbb{N}$ (time to live) is the maximum depth the message can reach. Note that advertisement $a$ is defined in the same space of queries, so PROSA relevance functions defined for queries can be used for advertisements as well.

When a peer wants to publicize a resource or a group of resources, it first builds the advertisement message. The message is then forwarded to a subset of neighbours, then to a set of second level neighbours, and so on until maximum depth $ttl$ is reached. More details can be found in [11].

III. Preliminary Results

A. The Data Set

The model we introduced in previous sections has several characteristics we did not notice in current real data sets, especially for what concerns the integration of trust within e-learning, however since different data sets (networks) within trust context and, separately, within e-learning are actually available, we exploit them by properly integrating their data in order to get to a data set with all features described in the proposed model. We preferred this choice rather than build a synthesized data set from scratch in order to perform simulations based on real data. Several data set have been used; in the following we briefly describe them.
Merlot [12] was our main source of information, with over 60000 users and 20000 resources. Merlot dataset contains information about difficulty level and also somehow about prerequisites/objectives (we described them using Merlot categories [13]). We then enhanced Merlot data set with data coming from other data sets; all these values are generated always in accordance with the statistical distribution of the data set they come from.

Ariadne[6] is an european project providing a distributed LOR accessible through WebServices. We used Ariadne to get information about the distribution of time spent by people while learning, in order to assign these values to resources. In addition to this, we adopt the heuristic that when associating time evaluation different peers give to the same resources, standard deviation is limited, reflecting the (realistic) case of avoiding scattered values of ”required time” to the same resource.

Epinions is a social network mainly used to figure out reputation [14]. Epinions data set was used to evaluate peer-resources trustworthiness. The heuristic we added is that trust is positive, modeling the realistic fact that a ”bad” resource is not inserted into the system. Note that Merlot allows to provide a comment about a resource, similar to a trustworthiness but too few values have been provided for this field.

Advogato[15] is an online community site dedicated to free software development, where users can certify each other on different levels (we name trustworthiness values). Advogato’s data set has been exploited to assign peer-to-peer trust values; no heuristic has been imposed for these values.

B. Simulation Results

A first set of simulation has been performed. In particular, we considered that Merlot data set presents some nodes (about 800 over 60000) being a sort of ”superpeers”, i.e. nodes pointing to many resources; this could be considered a reasonable scenario within e-learning context where few ”gurus” provide many resources, while many others access them. Due to this consideration, queries are not sent to all 60000 nodes, rather to a subset of about 10000 nodes (including superpeers and ”ordinary” nodes); note that the algorithm anyway operates over the entire set of 60000 nodes. In particular, we establish a set of significant query, i.e. whose preL and objL fall into the same Merlot category and are not directly related, so that a learning path is needed to link them; each of these 10 queries originates from each of 10000 peer of the network subset, and we count the number of peers that receive an answer for that query exploiting resources over 60000 nodes. Note that advertising mechanism has not been considered at this stage.

In fig. 1 we report on the left the number of answering nodes for each query (named with IDs); this measures the effectiveness of our algorithm over the data set, and these preliminary results are encouraging, since the worst query is q0 with about 70% (7000 over 10000) nodes with successful answer; others are better. On the right we show the capability of nodes of resolving queries; the main consideration we derive is that most nodes actually answers most queries (the max is 10), thus being a superpeer is not required, being even an ”ordinary” node able to answer queries.

IV. CONCLUSIONS

The work presented in this paper aimed at providing an effective e-learning system where users can get trusted and personalized resources, thanks to the underlying PROSA architecture. Preliminary results shown the effectiveness of this approach. Further work has though still be performed, in particular the proposed advertising mechanism must be fully implemented and simulated, in order to test its usefulness in discovering new hosted resources.

REFERENCES