LifeSocial.KOM: A Secure and P2P-based Solution for Online Social Networks

Kalman Graffi, Christian Gross, Dominik Stingl, Daniel Hartung, Aleksandra Kovacevic, and Ralf Steinmetz
Multimedia Communications Lab[1], Technische Universität Darmstadt, Germany
The Norwegian Information Security Laboratory[2], Gjøvik University College, Norway
Email: {graffi[1],cgross[1],stingl[1],sandra[1],steinmetz[1]}@kom.tu-darmstadt.de, daniel.hartung[2]@hig.no

***Abstract*— The phenomenon of online social networks reaches millions of users in the Internet nowadays. In these, users present themselves, their interests and their social links which they use to interact with other users. We present in this paper LifeSocial.KOM, a p2p-based platform for secure online social networks which provides the functionality of common online social networks in a totally distributed and secure manner. It is plugin-based, thus extendible in its functionality, providing secure communication and access-controlled storage as well as monitored quality of service, addressing the needs of both, users and system providers. The platform operates solely on the resources of the users, eliminating the concentration of crucial operational costs for one provider. In a testbed evaluation, we show the feasibility of the approach and point out the potential of the p2p paradigm in the field of online social networks.**

I. Introduction

The influence of Internet applications on daily life is continuously increasing. Starting as a pure communication platform for scientists, it revolutionized both the interaction between the people and the economy. Online social networks allow users to create profiles, link to their friends, publish photos and status updates and various forms of user-to-user interaction. Facebook has the largest user community with more than 450 Million profiles. Many communities exist, addressing user groups like students, researchers, musicians or businessmen.

As an alternative IT architecture to current platform architecture of common online social networks, we introduce in distributed approaches and propose a peer-to-peer (p2p) based platform for online social networks, distributing the load on the participating nodes. In Section II, we describe our solution pointing out the core (plugin-based) architecture and functional plugins. We also introduce the security concept and our approach to monitor the quality of the distributed architecture. In Section III, we present a brief evaluation showing the feasibility of the approach and the distribution of the load on the participating nodes. Section IV concludes with a brief summary and outlook on future work.

A. IT Architectures for Online Social Networks

Currently, most of the online social networks are operated by the massive usage of servers. On the one hand, several large databases are required to handle the profile and multimedia content of the users. On the other hand, web servers are needed to generate the page displayed to the user. As the current situation of online social networks shows, the server-based IT architecture provides the desired performance. The operational costs, however, are very high and the users do not trust platform providers regarding the security of their data.

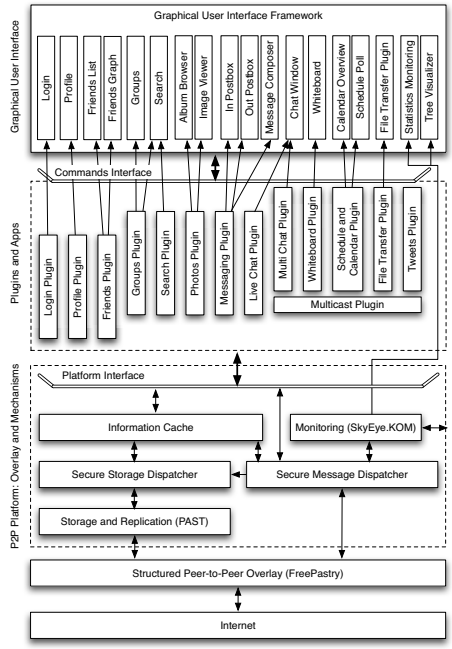
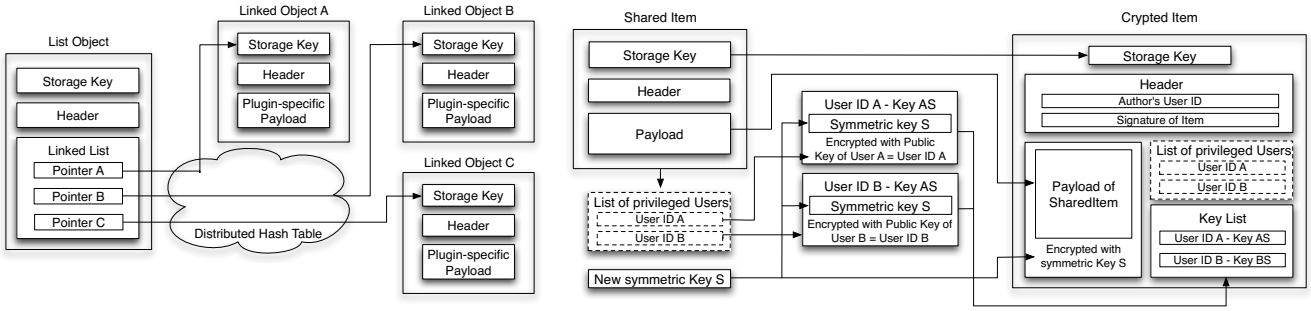


Fig. 1. Plugin-based Architecture of LifeSocial.KOM

Large scale networks for user interaction, however, also exist in other application fields. Skype, for example, has more than 21 Million users permanently online, had more than 1 Billion downloads up to Sept. 2008 and 450 Million users in the first quarter of 2009. This vast number of users is similar or even higher than the number of users in social networks. Skype shifts the maintenance and operational costs to the users by connecting them in a large-scale p2p network. Skype uses a globally decentralized user directory and provides user-to-user communication for free. The efficiency of p2p-based architectures has been demonstrated in various applications addressing tight quality requirements (Skype) and costly, high-traffic demands (file sharing and multimedia streaming).

We advocate that online social networks will be the next main application field for the p2p paradigm. In this paper, we present LifeSocial.KOM, a p2p-based online social network which demonstrates the feasibility of the distributed approach.

[1]This work has been partially funded by the DFG research group QuaP2P, EU SmoothIT and BMBF Premium Services



(a) Principle of a Distributed Linked List

(b) Distributed Access Control

Fig. 2. Data Management in LifeSocial.KOM

B. The Peer-to-Peer Paradigm

The p2p paradigm provides means for distributed platforms by totally distributing the load and the duties on the participating nodes. All functionality and responsibility in the system are provided by the cooperation of the nodes in a self-organizing approach. P2P overlays are networks on top of the Internet, introducing a new addressing scheme for the participating nodes. These are interconnected and form a network dedicated to a single, typically data-centric application. In the p2p overlay various kinds of data, e.g. profile information, friend lists or user photos, can be stored and retrieved.

Structured p2p overlay link the documents stored in the p2p overlay to responsible peers, based on the object IDs. Nodes which want to retrieve a special object route to the peer responsible for the object's ID and retrieve it. Both storing and retrieving objects generates a message overhead of $O(logN)$, resulting in a more efficient distributed storage solution. Maintaining the structure of mapping object IDs to peer IDs, however, requires periodic refreshing of the mapping status and thus is more endangered by node online dynamism. Our solution for a p2p-based online social network, LifeSocial.KOM, uses a structured p2p overlay for fault-tolerant and efficient data storage and builds the desired social functionality in a plugin-based manner on top. The whole application load is distributed over all participating nodes, while providing user-based access control with decentralized access control lists as well as data availability with a dedicated replication solution. Although the resulting performance is comparable to existing online social networks, no costs arise for the platform provider.

The research field of p2p-based online social networks is young and comprises only a few approaches. Safebook [1] creates a new p2p overlay for a secure online social network with specific rules for the topology. Although the idea is inspiring, the performance of the solution is still to be evaluated, the same counts for Persona [2]. Peerson [3] uses a structured p2p overlay es well, but eventually had issues with the chosen p2p overlay OpenDHT [4] and included server assistance. Cryptree [5] provides a file sharing network with social components. It resulted in the commercial application Wuala and uses a server for the key management and as data backup. Distributed online social networks, as described in [6], assume that users operate small web servers with their data, which is a strong assumption.

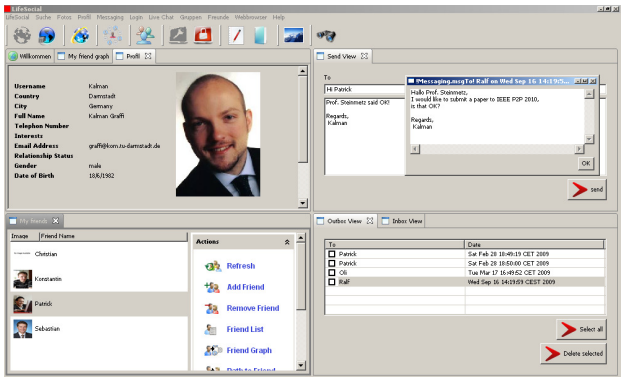
II. LifeSocial.KOM - A Distributed Platform for Secure Online Social Networks

The goal of LifeSocial.KOM is to provide a totally distributed, p2p-based online social network providing the functionality of common online social networks with additional user collaboration tools. As common functionality according to [7], online social networks offer user profiles, friend lists, user groups, photo albums, live chatting and status updates. In addition, we add the opportunity to exchange files, to collaborate in a group on a common whiteboard and to play online games, like Tic Tac Toe. Additional interaction functionality can be added as plugins, as shown in Figure 1, which are either based on already existing plugins, e.g. the multicast plugin, or create a functionality from the scratch. The resulting application's and the p2p-network's service quality are monitored, e.g. average data retrieval times are measured, allowing a provider to estimate the user quality experience.

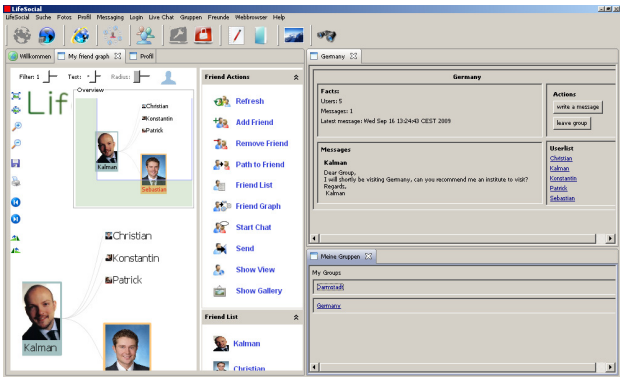
A. Core Peer-to-Peer Functionality

The core network layer of LifeSocial.KOM is a structured p2p overlay, providing the functionality for ID-based routing among peers. Several free p2p overlay networks exist, like JXTA, OpenDHT [4] or FreePastry [8]. The latter is based on Pastry [9] and provides a reliable storage component with integrated replication of the data called PAST [10]. With FreePastry and PAST objects can be very quickly and reliably stored and retrieved from the network based on their ID, with a lookup time of less then two seconds. For the scenario of online social network,s objects have an intrinsic ID, as they are always linked to a user or a group. For example the ID for the Profile object of user "Alice" could be "alice_profile". The main photo list of "Alice" may be "alice_photolist".

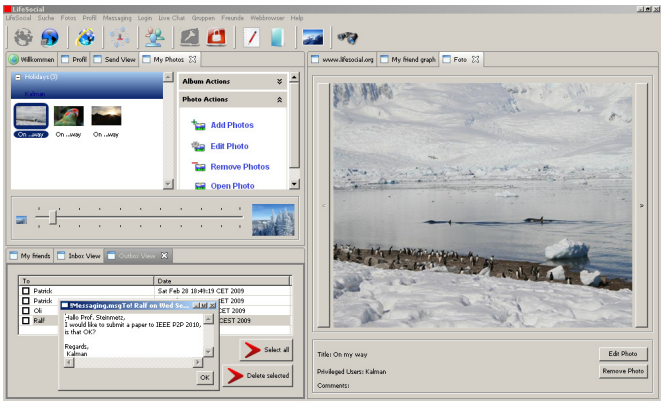
Data objects in LifeSocial.KOM contain either final information (e.g. a photo, a profile or a status entry) or additional links to further objects (e.g. a photo overview linking photo albums of the user). While final objects can be retrieved and instantly presented, objects with additional links result in a distributed linked list, as depicted in Figure 2(a), which can be traversed and retrieved recursively. One example for the distributed linked lists are user photos, which link to individual photo albums, which link to photos. Another example are friend lists (with IDs like "alice_friendlist") which contain the list of friends' profile IDs ("bob_profile", "charlie_profile").



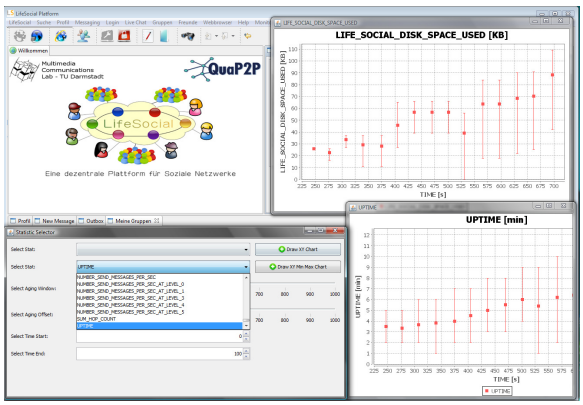
(a) Profile, Messaging and Friend List GUIs



(b) Friend Graph and Groups GUIs



(c) Photo Album and Messaging GUIs



(d) Monitoring View on System Quality

Fig. 3. Actual Screenshots of LifeSocial.KOM

Both final objects and distributed linked list objects are data objects that can be stored, replicated and efficiently retrieved in the structured p2p overlay.

In order to provide access control on the stored data, we apply our approach from [11] as depicted in Figure 2(b). We use cryptographic Public Keys as main user IDs in the network to uniquely identify users and to create an anchor for security, allowing the authentication of users, secure communication and user-based access control on the objects in the network. Any data object that is not public, is first encrypted with a symmetric cryptographic key. This symmetric key is encrypted individually with the Public Key of every read-enabled user (which are their user IDs) and appended to the data object. The list of encrypted symmetric keys as well as the object itself are signed by the author of the object and stored in the p2p overlay. Any other node interested in the object may retrieve it (the encrypted object) from the network and validate the signature using the Public Key of the author, but only the read-enabled users are able to decrypt the symmetric key and thus the content of the object. Concludingly, LifeSocial.KOM uses a data storage solution, which is totally distributed, shifting the load to the participating nodes, provides data availability through replication by PAST and user-based read/write access control. In addition, an *Information Cache* was added to keep the retrieved data objects in the cache for a while for further usage by the GUI and the functional plugins.

For messaging between peers, FreePastry provides a simple ID based routing of messages. In LifeSocial.KOM, we use a *Message Dispatcher* which dispatches incoming messages according to the message type to the internal plugins, e.g. assigning arriving chat updates to the chat plugin. We additionally encrypt messages with the Public Key (i.e. user ID) of the destination peer to secure the communication, the sender's authentication is provided by a signature to the message.

B. Plugin-based Application Composition

As main functional elements, we designed and implemented the following plugins for online social networks:

- Login: Registration / login based on cryptographic keys.
- Profile: Presenting a description of the user
- Friends: A list linking the profiles of the user's friends
- Messaging: Email-like exchange of messages
- Photo: A list of photo albums, linking to user photos
- Groups: A list of users joined in a common interest group
- Tweets: List of status updates of a user and its followers
- (Group) Chat: Direct user-to-user text messaging
- File transfer: Sending files from user to user
- Games for two with spectators: Tic Tac Toe (example)
- Whiteboard: Collaborative graphical editing of a canvas
- Schedule and Calendar: Collaborative time schedule
- Multicast: Creation and publishing to multicast groups

These functionalities are implemented in individual OSGi-based software components, called "plugins", which can be loaded and updated during runtime of the application. With the plugin-based architecture, LifeSocial.KOM can be extended easily. Every plugin comes with one or several (e.g. for

the Friends plugin) graphical user interfaces (GUI) that are arranged in a GUI framework, which allows for customizing the individual plugin windows. Four screenshots of the actual application are depicted in Figure 3. More screenshots can be found on www.lifesocial.org.

C. Monitoring Quality of Service

For the commercial application of the p2p paradigm, we need observable quality of service offered by the p2p application, which is hard to monitor in a totally distributed IT architecture. We addressed this issue in [12]. Our solution, called SkyEye.KOM, generates a statistical global view on an extendible list of various metrics, e.g. the object retrieval delays, storage, bandwidth and CPU consumption, number of nodes and various metrics more, over all peers. The statistical view comprises the global average, minimum and maximum values, as well as the sum, standard deviations and variances of the values. In order to keep the monitoring costs low, we create an easy to maintain tree topology among the peers in which every peer periodically sends its current status and the status of the sub-tree it manages to its parent node in the tree. The parent node aggregates the statistical information retrieved by calculating the averages, sums, counts and other statistical values. The size of the statistical representation does not increase with the size of the observed sample set. Every peer sends roughly the same amount of data (approx. 3 Kb) per update interval. The root of the tree, eventually, receives the statistics over the whole tree which comprises all of the peers. It calculates the global view on the statistics over all peers and pushes this information down the tree to all peers.

As every update message is answered with an acknowledgment message containing the global view, at the end, all peers are informed about the status of the network. The costs for providing this monitoring function are at about $2 \cdot (1+\beta) \cdot 3Kb$ per update interval (UI) with β being the degree of the tree. In an example with $\beta = 8$, $UI = 60s$ and $N = 1M$ peers, every peer has a bandwidth consumption of $54Kb/m$ or $0.9Kb/s$. The precision of the monitoring and the information age also depends on these three parameters (β, UI, N). The tree height is $O(log_\beta(N))$ and the monitoring age sum up to $2 \cdot O(log_\beta(N)) \cdot (update\ interval)$. In the same example, this results in a global view age of 6m 47s with $0.9Kb/s$ per peer.

Concludingly, the monitoring solution is very lightweight to operate, provides fresh results and helps the service provider to observe the quality of service in the distributed system. In Figure 3(d), we show the monitoring GUI on LifeSocial.KOM. In the lower left corner, the metric selector is depicted, which allows to pick among the monitored metrics one to present. Here, the uptime distribution of all peers and LifeSocial.KOM's storage space consumption are depicted in detail with the average value and the standard deviation.

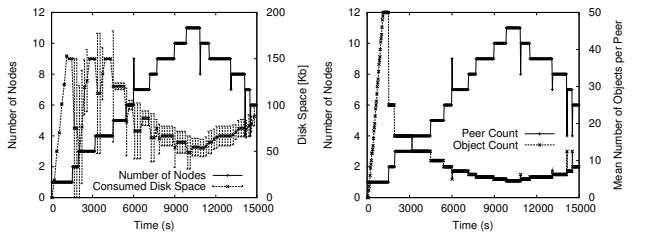
III. Evaluation and Testing

We implemented the p2p platform LifeSocial.KOM with integrated SkyEye.KOM as a Java-based standalone application since 2008. We implemented a set of plugins, providing the application of online social networks, as depicted in Figure 3, harnessing the resources of the participating peers to create a reliable p2p-based platform for social online networks. The data generated by the users is distributedly stored and replicated among the participating nodes and all load is deployed on the peers. This is shown in the first part of the evaluation. In the second part, we show the benefits of an integrated quality monitoring component, namely SkyEye.KOM, in a mid-sized enterprise environment. With a broad range of considered statistics in the p2p platform, we demonstrate the applicability of the proposed monitoring mechanisms in networks with variable node participation. In both scenarios we used SkyEye.KOM with $\beta = 2$ and an $update\_interval = 30s$.

In order to show the load and data distribution among the peers, we used a testbed with 11 peers which perform actions related to online social networks. First, 11 nodes joined one after another through a 15.000 seconds trial. The first node sets up his individual profile and friends list. Once all of the peers are online, 5 of the nodes leave the network.

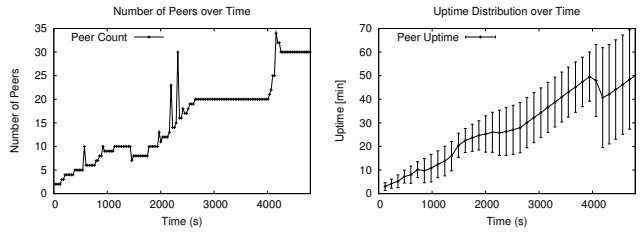
Figure 4(a) shows the mean disk usage in the network in relation to the number of peers in the p2p network over time. The mean disk usage per peer denotes that at the beginning the first peer created a set of objects which used 150 KB of disk space. With the second peer coming online, the objects are reassigned according to their object ID to the corresponding peer IDs. The standard deviation in the beginning (up to 4 nodes) is high, as every document is replicated in order to maintain a high object availability. The average number of objects per peer is depicted in Figure 4(b). Every object is replicated 4 times, which is reflected in the mean disk usage of 150 KB even with 4 nodes. With the fifth node joining, no further replication is needed and the mean disk usage drops. With every joining node, the load is shared among the peers, as depicted by the decreased mean disk utilization and the low standard deviation. As the peers start leaving the network, the distributed data objects need to be replicated again and the mean disk usage grows. The results show that the storage load is totally distributed among the participating nodes, while maintaining a high data availability.

In the second, enterprise sized testbed, we look at 30 peers which join subsequently in blocks of 10 peers. The first 10 peers join in the beginning in 20 minutes, create a profile and stabilize for 5 minutes. From $t = 1500s$ to $t = 2000s$ two of the peers left the network and joined again. Then the next 10 peers joined and the corresponding users created a profile. From $t = 3000s$ to $t = 3500s$ the peers request



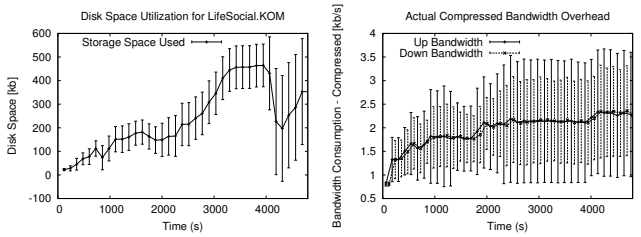
(a) Disk Usage (b) Peer and Object Count over Time

Fig. 4. Data Replication in LifeSocial.KOM in a Small-Scale Network



(a) Peer Count (b) Online Time Distribution

Fig. 5. Peer Count and Online Time in Enterprise-sized Testbed



(a) Disk Usage of LifeSocial.KOM (b) Compressed Bandwidth Usage

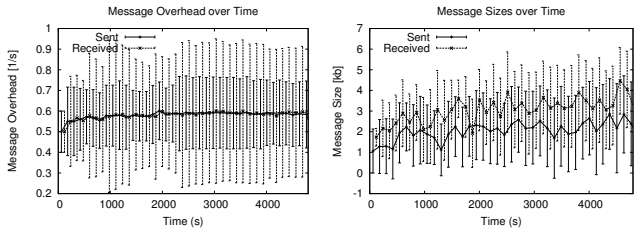
Fig. 6. Disk and Bandwidth Usage in Enterprise-sized Testbed

group and profile lookups. From $t = 4000s$ to $t = 4200s$, 10 more peers join. In this step of the evaluation we focus on giving a general view on the performance and costs of LifeSocial.KOM. In Figure 5(a), we visualize the peer count in the internal enterprise size network. The corresponding peer uptime distribution is depicted in Figure 5(b).

Regarding the overhead generated through LifeSocial, the main two metrics are the disk space usage and the bandwidth usage of LifeSocial.KOM. They are depicted in the Figures 6(a) and 6(b). While the disk space usage is around 150KB in the stabilization phase around $t = 1200s$, it grows in the following and stabilizes in the period from $t = 3000s$ on. Here, we observe the replication efforts that are done to maintain the data availability in LifeSocial.KOM. Traffic, being the scarcest resource, is compressed before transmitted. The bandwidth consumption is relatively small, LifeSocial.KOM needs only 2.1 to 2.5 KB/s both upload and download bandwidth. In Figure 7(a), we show the averaged total message overhead in the network, in Figure 7(b) the corresponding message sizes over the time. One observation is that the peer count has no influence on the message overhead, also the online time of the peers has no influence. All peers send up to 1 message per second and show a larger standard deviation for the received messages than for the sent ones. The total traffic in the network is depicted in Figure 6(b). We show that it is feasible and beneficial to integrate the monitoring component SkyEye.KOM in the p2p platform in order to enable users and platform providers to evaluate the service quality provided by the p2p system. The observed metrics in the evaluation show the flexibility of the approach to monitor metrics on various layers in the p2p system. Thus, a quality controlled and easy to extend general p2p platform for p2p applications has been built.

IV. Conclusion

Online social networks are very popular nowadays in the Internet. They allow users to create profiles and photo al-



(a) Messaging Overhead (b) Message Size Distribution

Fig. 7. Messaging Overhead and Size in Enterprise-sized Testbed

bums, link their friends and provide several communication and collaboration tools. In this paper, we presented LifeSocial.KOM, a p2p-based secure online social network, which shifts load for operating the infrastructure from the service providers to the users. LifeSocial.KOM uses FreePastry for interconnecting the participating nodes and PAST for reliable, replicated data storage. It has a plugin-based architecture on top, implementing the functionality of online social networks and some additional collaboration functionality. We extended the core architecture of LifeSocial.KOM with a security layer providing authenticated, secure communication and a user-based data access control. To monitor the service quality in the distributed architecture, we designed and implemented the monitoring solution SkyEye.KOM for structured p2p overlays. We presented a proof-of-concept of LifeSocial.KOM in a enterprise-sized testbed, showing the feasibility of the approach and the distribution of the load among the participating nodes. We believe that the p2p paradigm can drastically help online social network providers to cut their costs as well as that online social networks are the next big application area for p2p-based solutions. In the future, we plan to perform a public beta-test of LifeSocial.KOM in the Internet, in order to evaluate our solution in a large-scale live deployment.

References

[1] L. A. Cutillo, R. Molva, and T. Strufe, "Safebook : Feasibility of Transitive Cooperation for Privacy on a Decentralized Social Network," in *Proc. of IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications*, 2009.

[2] R. Baden *et al.*, "Persona: An Online Social Network with User-defined Privacy," in *Proc. of ACM SIGCOMM*, 2009.

[3] S. Buchegger *et al.*, "PeerSoN: P2P Social Networking - Early Experiences and Insights," in *Proc of. ACM Workshop on SNS*, 2009.

[4] S. Rhea *et al.*, "OpenDHT: A Public DHT Service and its Uses," in *Proc. of ACM SIGCOMM*, 2005.

[5] D. Grolimund *et al.*, "Cryptree: A Folder Tree Structure for Cryptographic File Systems," in *Proc. of SRDS*, 2006.

[6] C. Yeung *et al.*, "Decentralization: The Future of Online Social Networking," in *W3C Workshop on the Future of Social Networking Position Papers*, 2009.

[7] F. Benevenuto *et al.*, "Characterizing User Behavior in Online Social Networks," in *Proc. of ACM SIGCOMM*, 2009.

[8] Freepastry, http://www.freepastry.org/FreePastry/.

[9] A. I. T. Rowstron and P. Druschel, "Pastry: Scalable, Decentralized Object Location, and Routing for Large-Scale Peer-to-Peer Systems," in *Proc. of IFIP/ACM Middleware*, 2001.

[10] P. Druschel, "PAST: A Large-Scale, Persistent Peer-to-Peer Storage Utility," in *Proc. of HotOS*, 2001.

[11] K. Graffi *et al.*, "Practical Security in P2P-based Social Networks," in *Proc. of IEEE LCN*, 2009.

[12] K. Graffi *et al.*, "SkyEye.KOM: An Information Management Over-Overlay for Getting the Oracle View on Structured P2P Systems," in *Proc. of IEEE ICPADS*, 2008.