Proposition de recherche doctorale

Adaptive overlay technology for media distribution in the OSN context

Mots clés :

- **Directeur de thèse** : ARTUR HECKER
- **Co-encadrant(s)** :
- **Unité de recherche** : Laboratoire Traitement et Communication de l'Information
- **Ecole doctorale** : École Doctorale Informatique, Télécommunications, Électronique de Paris
- **Domaine scientifique principal**: Divers

Résumé du projet de recherche (Langue 1)

This thesis aims at studies of efficient media distribution architectures in the context of Online Social Networks (OSN). Recently, implementing OSNs as a distributed system has been proposed as a promising alternative to better preserve user privacy [1][2]. Indeed, just like security, privacy is majorly a question of control; yet, in the centralized environment the authority is necessarily with the OSN platform provider, who alas has non-negligible commercial incentives to exploit user data for other purposes. Specifically, the P2P technology has been used to tackle these issues: here, the OSN is implemented on top of a P2P system [3][4]. Several initiatives are ongoing, the most well-known being, from the academic perspective, Safebook [1] and PeerSON [3], and, from the open-source community, Diaspora [5]. Yet, beyond privacy issues, several additional questions deserve a deeper study. Today, the OSNs are used in different manners, and the usage is not limited to the mere message exchange. Owing to their communitarian nature, the OSNs increasingly become a way for content dissemination – and often not only for dissemination of links to the content stored elsewhere (e.g. to the videos on Youtube), but for the direct dissemination of the user-produced content like photos, videos, sounds, etc. Therefore, such content distribution and streaming capabilities need to be integrated in any modern OSN. That means that a given P2P system realization has to be prepared to efficiently deliver services very different in nature. In other words, the OSN communities and OSN developers should be offered possibilities to bend the underlying P2P technology to their respective needs. We explain that in more details. Indeed, while the environment, the required services, the available knowledge from the OSN, the participation in the P2P network (as of churn) will dynamically change in operation, the implementation of a participating peer stays the same. Therefore, the same running implementation participates in an actually constantly changing P2P system required to fulfil different services' demands. To do that, the peer will need to be capable of adapting to the new situation or needs, triggered to do so through some external stimuli (e.g. through a specific reconfiguration command from a content provider in an operated environment, or through a dynamic adaptation in an autonomous setup). A visible part of such adaptation will happen at the service interface: the respective QoS could be enhanced/degraded (e.g. video could be delivered in a lower resolution because a mobile user has moved into a worse connectivity zone), data security could be changed (e.g. encryption could be activated – in spite of battery considerations – because the new channel is not trustworthy), etc. Such adaptation can be, for most, considered locally by the peer, (pre-)specified by some policies. Yet, for us, there is a more intriguing question of the best general topology of the P2P network and of the relationship between the latter and: • the structure of the social graph; • media transport and delivery; • the systemic properties like resilience, robustness and security; • user privacy. The study of these relationships is the actual goal of this thesis. What is the best strategy for the P2P/OSN merger? What is the best topology for media delivery over a P2P OSN? Should the P2P topology directly follow the OSN contacts (darknets or F2F-like architecture, claimed to be better for privacy [6][7]), or should it rather be used as an independent communication middleware for the OSN (therefore increasing the degrees of freedom in communications, as e.g. in Skype)? Can different service requirements be fulfilled by the same structure? (e.g. should it be different for file transfer and delay-sensitive applications? or maybe even different for video and voice?) Besides, the overall system resilience – another study goal of this thesis – will also depend on the structural properties of the used P2P graph. Centralized P2P networks (i.e. P2P networks exhibiting high centrality) have very different properties than fully distributed networks, both in terms of routing (average path length is related to the network diameter, local routing table size is related to the node degree) and in terms of error robustness and security (e.g. disturbing nodes with high betweenness has a particularly high impact on the whole structure [8]). Finally, there seems to be a need for adaptation to the constraints of the underlying network. P2P topology adaptation to the underlying network in general and ISP/P2P collaborations in particular have received a lot of interest in the past [9][10][11]. More recent proposals include Oracle-like architectures, where the ISP essentially aids in the construction of network locality aware P2P structures [12]. While this seems like a very promising way to optimize data transport over the P2P overlay (to shorten delays in streaming, etc.), it results in a certain P2P topology, which however needs to be revisited in previously discussed terms. In conclusion, we can see that in search of a more resilient, distributed architecture, the P2P layer in the OSN context, at least conceptually, becomes an additional sub-layer between the underlying IP/ISP networks (underlay) and the web of human connections. Its articulation to the actual service needs, the OSN knowledge and the underlay constraints need to be studied and an optimal topology needs to be sketched in this thesis. However, given the partly contradicting needs (examples: a centralized topology would radically simplify the structure and minimize delays/transport efforts at the cost of error robustness and privacy; decentralized, hub-based – e.g. scale-free – topologies would yield almost constant routes and good error robustness by partly sacrificing privacy and becoming vulnerable to targeted attacks; uniformly distributed networks would exhibit good systemic resilience to external events, but transport costs, delays and route stability would become an issue, just as network control), the hope for a universally valid solution rapidly vanishes. Given the variety of needs, as a possible approach, we here propose to elaborate on the idea of adaptable P2P overlays, capable of dynamically rewiring [13] in runtime to a given global topology target by using only local node operations (a necessary constraint to guarantee scalability). The targeted topology could

depend on the current environmental situation (friendly/hostile environment, prevalent service, really used ISPs, etc). In our previous work, we have proposed such rewiring mechanisms for specific classes of P2P networks, most notably for DHT-based networks [14]. Departing from any typical, running DHT network (Chord, Pastry, Kademlia [15]) with its designed small-world properties, our mechanisms, executed on all peers, are not only backwards compatible (therefore easing deployment) but also able to produce either a uniform degree distribution or a power-law degree distribution (i.e. a scale-free network). Once the network is rewired, the routing over these topologies is also optimized to use the full existing knowledge [15]. Besides, these proposals enrich the classically key-based-search oriented DHT with blur complex queries support (through an optimal broadcast service that exploits the present structure). Such search facilities are necessary to support different OSN services. Finally, our previous work also considers integrating replication mechanisms with the existing P2P topology [16]. While the existing work shows the potential of such approaches and seems promising, the mechanisms proposed so far are only one first step towards the idea of a demand-driven, in-runtime reconfigurable P2P algorithm, capable of yielding anything from centralized over scale-free to uniformly distributed topologies. Such a generalization needs to be further studied in this thesis, since several questions are still open problems. Most notably, the thesis could address the questions of triggers/indicators for topology reconfiguration, peer incentives and payback, related to becoming super-non a peer commitment/refusal protocol to such proposals. Also, while we used the DHTs as a departure point, inverse approaches might be interesting as well (e.g. random networks). Most importantly, in our previous work, we did not consider at all any available OSN knowledge, which could provide further hints towards a currently optimal topology.

Références [1] L.A. Cutillo, R.Molva, T.Strufe, “Safebook: a privacy preserving online social network leveraging on real-life trust”, IEEE Communications Magazine, vol 47, No 12, Consumer Communications and Networking Series, December 2009. [2] A. Shakimov, A. Varshavsky, L.P. Cox, R.Caceres, “Privacy, cost, and availability tradeoffs in decentralized OSNs”, in proceedings of the 2nd ACM workshop on Online Social Networks, ser. WOSN ’09, 2009, pp 13-18. [3] Sonja Buchegger, Doris Schiöberg, Le Hung Vu, Anwitaman Datta. PeerSoN: P2P Social Networking - Early Experiences and Insights. In Proceedings of SocialNets 2009, The 2nd Workshop on Social Network Systems, Nuernberg, Germany, March 31, 2009. pdf bib [4] Sonja Buchegger, Anwitaman Datta. A Case for P2P Infrastructure for Social Networks - Opportunities and Challenges. In Proceedings of WONS 2009, The Sixth International Conference on Wireless On-demand Network Systems and Services, Snowbird, Utah, USA, February 2-4 2009. [5] Diaspora P2P OSN, https://joindiaspora.com/ [6] B. C. Popescu, B. Crispo, A. S. Tannenbaum, "Safe and Private Data Sharing with Turtle: Friends Team-Up and Beat the System" Presented at the 12th Cambridge International Workshop on Security Protocols, April 2004 [7] I. Clarke, O. Sandberg, B. Wiley, T. W. Hong, “Freenet: A Distributed Anonymous Information Storage and Retrieval System”, Lecture notes in computer science, Vol. 2009, pp. 46-66, Springer, 2001 (https://freenetproject.org/). [8] R. Albert, H. Jeong, A.L.Barabasi, “Error and Attack Tolerance of Complex Networks”, Nature, vol. 406, 2000. [9] A. Nakao, L. Peterson, and A. Bavier, “A Routing Underlay for Overlay Networks,” in SIGCOMM, 2003. [10] S. Seetharaman and M. Ammar, “On the Interaction between Dynamic Routing in the Native and Overlay Layers,” in INFOCOM 2006. [11] S. Ratnasamy, M. Handley, R. Karp, and S. Shenker, “Topologically aware overlay construction and server selection,” in INFOCOM 2002. [12] V. Aggarwal, A. Feldmann, C. Scheideler, “Can ISPs and P2P systems co-operate for improved performance?” ACM SIGCOMM Computer Communications Review (CCR), 37(3):29–40, July 2007. [13] S. Ktari, A. Hecker, H. Labiod, "A Construction Scheme for Scale-free DHT-based Networks", IEEE GLOBECOM, December 2009, Hawaii USA. [14] S. Ktari, A. Hecker, H. Labiod, “Power-Law Chord Architecture in P2P Overlays”, ACM CoNext, Madrid, Spain, December 2008. [15] S. Ktari, A. Hecker, H. Labiod, “Symmetric routing in DHT overlays”, Telecommunication Systems Journal, Nr. 1018 4864, DOI: 10.1007/s11235-010 9326-y, Juin 2010. [16] S. Ktari, M. Zoubert, A. Hecker and H. Labiod, “Symmetric Replication for Efficient Flooding in DHTs”, ACM MobiHoc 2008, Hong Kong, China.