Compliance Management in Peer-to-Peer Networks

Alexander Schneider, *Institut für Informatik, Heinrich Heine Universität Düsseldorf*

***Abstract*—Compliance management for peer-to-peer networks describes a process ensuring that content inside the network is distributed and stored in a way that does not violate user-defined preferences. To our knowledge there are no existing architectures which allow for compliance management in peer-to-peer networks. Our research goal is to create network-structures which enable and facilitate compliance management. We already created a proof-of-concept, which was evaluated in a simulator and shows that the concept of compliance management is feasible.**

***Index Terms*—Netsys, extended abstract, PhD Forum, Compliance Management**

I. Introduction

USUALLY participants in peer-to-peer networks, which distribute content, need to contribute to the health of the network by sharing and storing content. This may cause legal or ethical problems with some of the content. Compliance management aims to solve this problem, by allowing every user to define general categories of content she is willing to store and forward. Hence the user can participate without fear of breaking the law or handling unwanted content. Given a chunk of data that is described by a list of tags, we seek to answer the following question: is it possible to store and forward it in such a way, that none of the individual preferences of the users are violated?

A typical use-case for compliance management would be a file-sharing network, where the user can participate without perpetrating files that are illegal in her country. Another use case could be social networks that use compliance management inside a network comprised solely of company-owned servers in geographically distinct locations. Often social networks have one guideline for appropriate content for all users. With compliance management they could serve different content depending on the origin of the user automatically, thus enabling the social network to serve acceptable content according to social norms of the users, instead of trying to have a general ruleset which tries to fit all users cultural and social preferences.

In order to specify which content is acceptable for a given node, we propose that each chunk of data handled by the peer-to-peer network is assigned one or more tags describing its content. We acknowledge that assigning these tags in a reliable and trustworthy way is a significant challenge that we currently do not address at length. We will, however, provide reasoning why we believe this to be a solvable problem.

II. Current Approach

The main challenge of our research topic is to design a peer-to-peer architecture, which performs well under the constraint that every user only cooperates for a fraction of the data, which the network is supposed to contain.

Author-Mail: aschneider@cs.uni-duesseldorf.de

A. Prototype Architecture

As a first step towards such an architecture we proposed a system, where all data is assigned tags, according to some classification. All participating nodes have to announce their policies, which are a list of tags the participant is willing to store and forward. A multi-tiered routing table is used to find suitable known contacts which can forward some data or data-request. This system is limited in the regard that it can only handle a finite number of tags in the system, which are defined as a bootstrap parameter.

To show that this architecture is feasible, we modified the Kademlia peer-to-peer overlay to use tags, node-policies and multi-tiered routing tables. We ran several simulations using this Kademlia approach and showed that in general the network was feasible. However, we only simulated with no churn in mind.

We implemented the modified Kademlia overlay, which we called Comademlia, inside the PeerfactSIM.KOM simulator [2]. We tried to accomplish high content retrieval rates through redundant storage inside the network. In our simulations we achieved retrieval rates of 100% for most simulations and near perfect retrieval rates for the rest. Since we used a multi-tiered routing structure this was achieved with a trade-off for the state-complexity for a participating node, meaning that a node had to store more information than in a vanilla Kademlia network.

B. Tagging

We assume that all chunks of content are associated with tags, that characterize the content. Tags either describe the content, e.g. violence or explicit speech, or they provide meta information such as legal for all audiences in Ger- many. They are assigned by trusted parties or by means of collective decisions. The latter is very similar to what is regularly done in order to realize quality control at web-sites such as stackoverflow.com. Each node in the peer-to-peer network specifies its policy by maintaining a list of tags. A node will not participate in routing and storing content, with tags that are not contained in its policy. Of course the connection between a tag and some data has to be trustworthy, which can be achieved by e.g. cryptographic signatures. A good and efficient system for the assignment of tags - in particular in form of collective decisions - is certainly an interesting research challenge. However, given that content classification is regularly done in other contexts both in a centralized fashion

and as collective decisions, leads us to the assumption that developing such a system is generally feasible and is thus one of our future challenges.

C. Bloom Filters

Another approach we are researching is utilizing (aggregated) bloom-filters for the propagation of content-policy information of network-nodes. This could lead to more elegant routing algorithms, and also removes the constraint of a finite number of tags. Every node would insert their accepted tags in a bloom filter and announce the bloom filter to all known neighbors. The neighbors then use the bloom filters of their neighbors to construct aggregated bloom filters which they in turn again can propagate to their neighbors. Repeating this procedure could enable network nodes to posses aggregated bloom filters for all neighbors with hop-distance n. Furthermore, since bloom-filters are only a collection of fitting hashing algorithms, the inputs are not limited and thus any number of different tags can be used.

D. Challenges

A compliance management system has several challenges to solve. For one there is the theoretical possibility, that any node has only neighbors which do not handle some desired content. Preventing or mitigating such an ”eclipse” is crucial for every compliance management network.
As with all peer-to-peer networks, churn can be a destructive factor as well. Churn is especially vicious against compliance management networks, since data can not be replicated arbitrarily, but only at certain network nodes, that accept it.

III. Further Research

Big online content provider usually do not use peer-to-peer structures, but instead employ methods like content-delivery networks (CDNs). We want to examine whether there is the possibility of combining a peer-to-peer approach with current generation CDNs to provide compliance management.

Another approach we are aiming to explore is to use named data networking [1] as an overlay. Named data networking routes on data-names instead of e.g. IP addresses. Our idea is to use tags as name-prefixes to enable compliant routing on arbitrary network topologies.

Lastly, it would be interesting to know if the topology of a network have any influence over the performance of compliance management or whether the network topology is irrelevant.

IV. Conclusion

Our research is exploring methods to introduce networks, which use compliance management as a means of giving network participants a granular way of filtering data that is handled by them. A functioning, performant compliance management peer-to-peer network could contribute to a new wave of popular peer-to-peer networks, since the handling of unwanted content is one of the unappealing aspects of some of the current peer-to-peer networks.

References

[1] L. Zhang, D. Estrin, J. Burke, V. Jacobson, J. D. Thornton, D. K. Smetters, B. Zhang, G. Tsudik, D. Massey, C. Papadopoulos *et al.*, “Named data networking (ndn) project,” *Relatório Técnico NDN-0001, Xerox Palo Alto Research Center-PARC*, 2010.

[2] D. Stingl, C. Gross, J. Rückert, L. Nobach, A. Kovacevic, and R. Steinmetz, “Peerfactsim. kom: A simulation framework for peer-to-peer systems,” in *High Performance Computing and Simulation (HPCS), 2011 International Conference on*. IEEE, 2011, pp. 577–584.