WhatsUp: news from, for, through everyone

Antoine Boutet, Davide Frey, Rachid Guerraoui, Anne-Marie Kermarrec

WhatsUp: News, From, For, Through, Everyone

A. Boutet,[1] D. Frey[1], R. Guerraoui[2] & A.-M. Kermarrec[1]
[1] INRIA Rennes Bretagne-Atlantique - [2] EPFL

***Abstract*—WHATSUP (WUP) is a new form of electronic news. It is personalized and decentralized. Users receive news and have the ability to express their interest in it. This opinion, in turn, is used as an implicit and dynamic subscription scheme to filter and personalize future information. The system is peer-to-peer: no big brother company controls the news, and no central server makes it vulnerable to failures, censorship or attacks.**

At the heart of WUP lies the idea of collaborative filtering applied to the dissemination of news: people who liked the same news in the past might as well like the same news in the future: irrelevant news disappear by themselves. The idea is put to work through BEEP: a biased epidemic dissemination (gossip) protocol that delivers news to interested users in a timely manner, despite jamming and churn. BEEP is dynamically parameterized on a per-user, per-news, and per-dissemination-hop basis. When compared to a classical epidemic dissemination protocol, BEEP has two key characteristics: *orientation* and *amplification*. Every user forwards the news of interest to a randomly selected set of users largely constituted by those who have similar interests (*orientation*). Moreover, the size of this set of users depends on the level of interest in the news itself (*amplification*).

***Index Terms*—Rumor spreading, network overlay, gossip, social networks, decentralized system, electronic news.**

I. PITCH

Unlike its predecessor, the Web 2.0 is not a read-only infrastructure but a collaborative read-write platform with active players. Users classify Web content based on their interests, and share it with their friends or even unknown users. Users also post blogs as well as real-time news.

Web 2.0 applications, such as LastFM, Flickr, CiteUlike, Delicious, Twitter or Facebook, gather hundreds of millions of users and play a major role in worldwide information dissemination. Recent events in Iran and Haïti highlighted the crucial importance of this emergent medium [1]. Nonetheless, the large amounts of data generated by users often makes it hard to determine which content is relevant to whom. It is common to overcome this problem using recommendation technologies such as *collaborative* or *social* filtering. Collaborative filtering consists in gathering users who expressed similar interests and relies on the resulting overlays to generate future recommendations. On the other hand, social filtering suggests items to a user which have been marked as interesting in her neighborhood.

So far, collaborative filtering has mainly been applied to systems, such as Amazon, or Netflix, that recommend music, books, or movies to their users. On the other hand, news dissemination, and in general Web 2.0 sites targeting highly dynamic content, such as Digg, have been relying mainly on social filters to identify the most interesting articles, pictures or movies submitted.

We believe that relying on explicitly declared friends to obtain new information strongly limits the content that can be received [2]. Declaring someone as friend one day does by no means qualify the news posted by that "friend" as highly relevant a few months later (Facebook provides anecdotal evidence of that). Rather, relevant and important news often come from people outside one's explicit circle of friends.

In addition, most Web 2.0 applications have been hosted by few servers and managed by a single company. While this is explained by computationally intensive approaches that only large companies can afford, this creates obvious vulnerabilities. A government might pressure a company to stop its activities in certain areas or might filter some piece of information or the whole system [3]. Furthermore a system hosted by a set of servers can get attacked by some form of news bombing or meet infrastructure failures or scalability issues [4].

In short, WHATSUP [5] aims to address these limitations by relying on collaborative filtering to disseminate personalized news, in a completely decentralized manner.

II. OVERVIEW OF WUP

WHATSUP users express their interests simply by saying whether they like each piece of news they read. WHATSUP (WUP), in turn, transforms these interests into implicit subscriptions and leverages these to drive the dissemination of news to interested parties. This establishes a feedback mechanism by which the more opinionated a user, the more personalized the information she will get and the faster she will get it.

The decentralized architecture of WUP makes it very hard for governments or any big-brother companies to filter the news. Additionally, its information dissemination process is resistant to failures of computer and networking infrastructures. WUP's personalization mechanism prevents users from being spammed with irrelevant news, while allowing them to receive interesting ones in a split second.

We base WUP on a dynamic overlay of implicit acquaintances build through a collaborative-filtering system. A relationship is established when two users have expressed similar interests in the same news. The dissemination of information is then performed using a novel biased epidemic (gossip) protocol (BEEP).

BEEP improves on existing epidemic protocols in several ways. While epidemic protocols are typically known to be (a) simple to deploy and (b) resistant to dynamics, they cannot be used directly to disseminate news in the real world. Their inherent uniformity causes all news and all users to be treated

equally. In practice, they tend to deliver all news to everyone. Also, in their simple form, these protocols are quite vulnerable to attacks. Malicious users can jam or overload the system with bogus news or news with non-acceptable content, scaring away potential users and preventing them from using and contributing to the system.

BEEP addresses these drawbacks through two main ideas: *amplification* and *orientation*. In a nutshell, BEEP parameterizes the dissemination protocol on a per-user/per-news basis rather than applying the same dissemination parameters to the whole system. *Amplification* means that the size of the set of nodes to which a node forwards some information depends on the interest expressed by the user at that node. Clearly, *amplification* is used to act as a social filter, depending on the opinion of many users, the news can (a) die in case of irrelevance, (b) be well propagated, or (c) make buzz. The goal here is to identify the importance of the news in order to adapt its propagation through the network. The *amplification* process also takes into account parameters such as the origin of the news. If some information comes from nodes that are not well connected in the network (and possibly part of the same malicious coalition), the information is more likely to be true. Moreover, taking this into account makes it possible to avoid the bias in the dissemination introduced by colluding nodes as observed in Digg [6].

The goal of *orientation* is instead to bias the constitution of the set of nodes to which a user transmits the news at her disposal. This set largely consists of the implicit acquaintances of a node built through collaborative filtering, complemented by a subset of nodes chosen randomly. In addition, *orientation* can take into account parameters such as bandwidth restrictions in order to adapt the traffic to and from nodes according to their network capacities.

III. DEMONSTRATION

Our demonstration will illustrate the advantages of collaborative filtering in WUP, and contrast them with the drawbacks of news dissemination through the explicit friends of a social network. We will achieve this by comparing WUP's dissemination against that achieved by existing systems such as Digg and Twitter.

Digg is a social network which aims to discover and share content. The value of content is collectively determined: once something is submitted, other people see it and "Digg" what they like best. If a submission receives enough Diggs, it is promoted as popular and benefits from better visibility. Twitter is an emergent micro-blogging service that enables users to broadcast and share information about their statuses, opinions and activities within a limit of 140 characters per message. Due to the different nature of these social networks, Twitter has no social voting (as in Digg) to allow the users express their opinions about the news they receive. However, in terms of rumor spreading, the retweet mechanism achieves a similar purpose. A user who retweets a piece of information (forwards it to all her followers) can be viewed as interested in it.

Our demonstration will first use a visualization tool to depict the evolution of the dissemination through Digg, and Twitter. The tool will also evaluate news spreading in these systems in terms of (a) the lack of similarity of interests between explicit neighbors, and (b) the amount of spam (non-interesting news received) and duplicates.

Second, it will demonstrate WUP using both simulations based on the traces from Digg and Twitter, and a PlanetLab deployment. The simulations will show the effectiveness of WUP in disseminating news, while the deployment will allow conference attendees to interact with the system by rating news and benefiting from WUP's dissemination mechanism. Specifically, users will be able to drop rumors into the system and express their opinions about the news they receive. The PlanetLab nodes will act as real users, their behaviors being driven by our collected traces. As in the case of Digg and Twitter, we will use our visualization tool to display the status of the system, the topology and the dynamics of the network, and the location of the attendees in the system. In addition, the deployment will make it possible to measure various network metrics, such as the number of messages to be disseminated and the bandwidth consumed.

IV. SUMMARY

We propose WHATSUP, a decentralized and personalized system based on an epidemic dissemination (gossip) protocol using adaptive fanout to disseminate information in a fast, personalized and free manner. WHATSUP aims to be resistant to censorship and infrastructure failures.

Our demonstration will consist of two phases. We will highlight the limitation of explicit friends to recommend news and then show how WHATSUP leverages implicit friends in a more adequate manner.

We will present our evaluation of WHATSUP through simulation and a real deployment. Ultimately, our aim is to enable conference attendees to visualize and to interact with the system.

REFERENCES

[1] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" in *WWW '10: Proceedings of the 19th international conference on World wide web*. New York, NY, USA: ACM, 2010, pp. 591–600.

[2] F. Wu, B. A. Huberman, L. A. Adamic, and J. R. Tyler, "Information flow in social groups," *Physica A: Statistical and Theoretical Physics*, vol. 337, no. 1-2, pp. 327 – 335, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/B6TVG-4BNMC3W-2/2/647d610316bf69d693bc28ba370ae5af

[3] http://en.wikipedia.org/wiki/Internet_censorship_in_Iran, "Internet censorship in iran."

[4] http://twitter.com/503, "Twitter is over capacity."

[5] http://www.gossple.fr/whatsup, "Whatsup: News, from, for, through, everyone."

[6] K. Lerman, "Social networks and social information filtering on digg," in *Proceedings of 1st International Conference on Weblogs and Social Media (ICWSM-07)*, 2007.