Future Generation Computer Systems 86 (2018) 775–791

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Efficient service discovery in decentralized online social networks

Bo Yuan [a,b], Lu Liu [b,*], Nick Antonopoulos [b]

[a] The Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai, China
[b] College of Engineering and Technology, University of Derby, Derby DE22 1GB, United Kingdom

HIGHLIGHTS

- A homophily-based user model based on social relationship and semantic content.
- An adaptive olfactory-sensitive search algorithm.
- Improved service discovery performance and reduced network traffic.

ARTICLE INFO

Article history:
Received 15 November 2016
Received in revised form 1 April 2017
Accepted 13 April 2017
Available online 19 May 2017

Keywords:
Service discovery
Decentralized online social networks
Peer-to-Peer
Swarm intelligence

ABSTRACT

Online social networks (OSN) have attracted millions of users worldwide over the last decade. There are a series of urgent issues faced by existing OSN such as information overload, single-point of failure and privacy concerns. The booming Internet of Things (IoT) and Cloud Computing provide paradigms for the development of decentralized OSN. In this paper, we build a self-organized decentralized OSN (SDOSN) on the overlay network of an IoT infrastructure resembling real life social graph. A user model based on homophily features is proposed considering social relationships and user interests and focuses on the key OSN functionality of efficient information dissemination. A swarm intelligence search method is also proposed to facilitate adaptive forwarding and effective service discovery. Our evaluation, performed in simulation using real-world datasets, shows that our approach achieves better performance when compared with the state-of-the-art methods in a dynamic network environment.

1. Introduction

Online social networks (OSN) connect millions of Internet users and have played a principal role in changing the world into a new era of globalization, i.e. the online society. OSN websites, such as Facebook, Myspace, Twitter, Google+, and Flickr have become increasingly popular online applications for people to communicate information, share moments, and distribute ideas. As of 2016, the number of social network users reached 2.3 billion, which represents 68.3 percent of Internet users, and these figures are expected to increase in future years [1]. Most OSNs sites provide free storage for users to share their social content. In 2014, the data warehouse of Facebook, the largest social website, stored upwards of 300 PB of Hive data using a cloud system, with an incoming daily rate of about 600 TB user-generated content [2]. Social data is being produced faster than any organization has had to deal with before. To add to the challenge, the data is heterogeneous and produced in many formats, including plain text, document, image, video and so on. This flood of data is being generated from any number of connected devices—from PCs, smart phones, tablets to streaming set-top boxes, gaming consoles, digital cameras and even fitness sensors. It is evident that we are entering into a social big data world that is urging the development of novel computing models and new types of architecture to cope with the sheer volume of data. Cloud Computing is a model for enabling ubiquitous, on-demand access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort [3]. After nearly ten years' development, Cloud Computing is maturing and addressing barriers for big data problems. Therefore, OSN can be hosted on Cloud platforms to utilize scalable computing and storage services to process their data in either privately owned, or third-party data centers with lower infrastructure costs [4].

However, all the mainstream OSN providers have a cloud based infrastructure that is designed with a logically centralized architecture controlled by a single authority, i.e. the social network service provider. Though large websites use content distribution networks and distributed computing for performance reasons, there is still a central repository for user data. In this sense, all sensitive personal data is stored in the providers' repository and so users are susceptible to loss of control over ownership of their privacy.

* Corresponding author.
E-mail address: L.Liu@derby.ac.uk (L. Liu).

http://dx.doi.org/10.1016/j.future.2017.04.022
0167-739X/

Furthermore, along with the tremendous amount of user requests and social data, the performance bottleneck has become a critical problem in real-time systems to ensure a smooth continuation and acceptable performance. Besides, the centralized nature has other drawbacks including single points of failure, a need to be online for every transaction and a lack of locality. These issues have provided momentum to the research and development of open, decentralized alternatives; this provides the motivation and context for our work. The architecture of decentralized OSN (DOSN) is created through participation by a set of autonomous OSN users collaborating with each other without a centralized repository. The social data will be stored in local devices and controlled by end users instead of OSN providers. In this way, the huge amount of social data will be distributed in potentially hundreds of millions of end devices rather than a single monolithic system. Therefore, DOSN can alleviate the rigid privacy-control issues, as well as provide adequate flexibility and capability to deal with big data problems

In the meanwhile, the mushroom growth of the IoT and the rapid development of associated technologies provide paradigms for the development of new DOSN. The IoT is a technical revolution that leads the development of next-generation of computing and communications. The IoT can utilize high-end computing processing power, provide interoperable networks and communication protocols and achieve greater flexibility and availability. Moreover, the Social IoT (SIoT) has become an emerging trend of augmenting physical devices and objects with social capability in order to make full use of the collective resources of connected "things". In any SIoT scenario, entities connect to form social groups, enabling collaboration to achieve specific collective goals. It is estimated that the IoT will consist of almost 50 billion objects by 2020 [5]. This prediction suggests that it is reasonable to anticipate that "things" will also form enormous social networks: thing-to-thing social networks resembling human society in the near future.

The infrastructure of the thing-to-thing social networks can be generalized as Peer-to-Peer (P2P) social networks. In recent years, P2P networks [6–11] have been proposed to support the decentralized infrastructure of OSN. The similarity between P2P networks and social networks, where peers can be considered as Internet users and connections reflect social relationships, leads us to believe the principles of P2P networks are suitable to guide the research on the design of DOSN. The nature of P2P networks resembles a real-life social graph built from bottom to top rather than the monolithic, top-down centralized structure that has existed to date, providing people with more control over what they share. Moreover, it possesses many favorable topological features such as self-organization, strong adaptability and fine scalability for large-scale networking applications. Furthermore, P2P networks can gather and harness tremendous computation and storage across the Internet. With SIoT technology, intelligent devices, vehicles, buildings and other objects embedded with sensing, computing, and communication capabilities can participate in an OSN with a dual role of computational processing as well as the content provision.

Until recently, P2P networks have been principally used for file sharing applications. Research into the mechanisms of DOSN, which is still at an early stage, has identified many problems and challenges, which are yet to be solved [9,10]. We examine two major research problems, namely how to design the architecture of DOSN and how to support efficient social information dissemination and service discovery. It should be noted that the service mentioned in the paper is a general concept that is not restricted to well-defined web services, but many types of data including text, hashtags, pictures or video clips in social networking systems. The service discovery should be considered to be a capability to locate varied, multisource social resources.

This paper proposes a self-organized decentralized social network (SDOSN) featured by P2P infrastructure with a swarm intelligence [12] search strategy. Specifically, a self-organized P2P social overlay network is introduced to form a local knowledge index that can facilitate accurate and personalized service discovery. The key challenges of this stage are how to define similarity measurements for social users to establish acquaintance shortcuts and how to utilize limited knowledge to make predictions for neighbor selection. A homophily-based user model is proposed to select promising neighbors in each routing step considering social relationships as well as content semantic features. Furthermore, a novel olfactory sensitive search (OSS) algorithm is proposed by exploring the swarm intelligence. The free and uncertain foraging behavior of swarm collaboration is abstracted and modeled to optimize the service discovery process in the DOSN. The OSS is well suited to this environment because it integrates knowledge of collective intelligence and the highly self-organized social features of users. In this way, a user can progressively gain experience during the discovery process and make future discovery more focused and accurate.

The performance of the proposed approach is evaluated using simulation experiments. A P2P social network platform is developed that is able to simulate network structure dynamisms (topology construction, churns, and routing) and service discovery processes (indexing, bootstrapping, and searching). The experiments use real-world datasets in three different network structures. The results show that the performance of the proposed service discovery algorithm in SDOSN achieves less average visited nodes, higher success rate and a higher recall when compared against existing state-of-the-art methods.

The main contributions of this paper are summarized as follows:

- A self-organized architecture for social overlay network of DOSN is proposed. This architecture fills the research gap between the traditional P2P infrastructure and the decentralized architecture of OSN.
- A homophily-based user model is introduced to capture the homophily similarity that integrates social relationship and user interest. This model is able to identify promising neighbors those are similar to the service provider and have the high number of connections.
- A novel olfactory sensitive search algorithm is proposed to guide the service discovery with an adaptive forwarding degree. The algorithm utilizes the collective swarm intelligence to discover the shortest paths with maximum desired services.
- A software simulation platform is designed and developed. It can simulate dynamic unstructured P2P networks with configurable routing protocols to support social overlay networks, decentralized service discovery, and evaluation of search models.

The remainder of the paper is structured as follows. Related works are identified in Section 2. Discussions on the design of self-organized architecture and supporting service discovery mechanisms are presented in Section 3. The homophily-based user model and the matchmaking method are presented in Section 4. Section 5 introduces the swarm intelligence algorithm necessary to achieve adaptive forwarding. The simulation platform, experiments and results are discussed in Section 6, and finally the conclusion is presented in Section 7.

2. Related works

In this section, the related works of service discovery in decentralized architectures are compared and contrasted. In decentralized architectures, all nodes are considered to be equal and an