Caching Mechanism of Decentralized Social Networks

Chenle Xiong

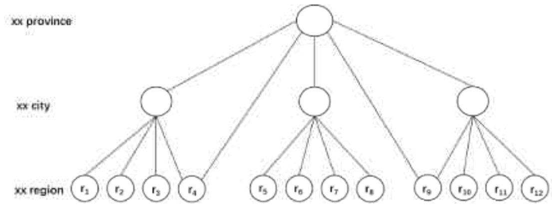
***Abstract*—With the rapid development of the Internet, the social traffic in the network is growing explosively. The cache pressure faced by social networks is increasing. However, traditional social networks are centralized. The cache content in the network is stored in the central server or base station, while the cache content of decentralized social networks is scattered in each node of the network. Therefore, this paper analyzes the cache mechanism and cache replacement strategy of decentralized social networks, which are completely different from centralized social networks.**

***Index Terms*—Decentralization, social network, seed file, caching mechanism.**

I. Introduction

The system model in decentralized social networks is composed of all user devices[1], and mobile devices communicate through P2P channels. Wherein, the set of users in the network is recorded as U={u1, u2,..., un}. Assuming that all users in the social network have communication coverage in region R, the communication coverage of region R is divided into three levels according to the actual provinces and cities, such as xx district, xx city, xx province. Mark all regions in the network as R={r1, r2,..., rm}. All regions form a tree structure. As shown in Figure 1 below, the region where a user u1 is located is marked as ru1, and ru1 is marked with the province ID, city ID, and region ID of the user u1.

TABLE I. AREA CLASSIFICATION MAP



In the whole network, we record the cache file as F, and the cache file set as F={f1, f2,... fq}, where f1=f2=...=fq. To ensure that the file is not lost, the cache file is cached by q different users, assuming that the size of each cache file is the same. If the cache file size is different, it can also be considered as a cache file that has been split into the same size. The cache space size of each user device in the network is different. The cache space capacity of the user device is recorded as Vu.

II. Cache Mechanism

Most cache strategies are designed based on the distance factor[2]. We have improved the design and proposed a cache strategy that comprehensively considers the distance, activity, contribution value and stability. When a file needs to be cached by users, which users should be given the file to cache, which users should be given priority, and which users should be given the file to cache. We will define the three-level cache list according to four factors. The cache file is first handed over to the user cache in the first level list, then to the user cache in the second level list, and finally to the user cache in the third level list.

We grade the cache list according to the priority distance, and then grade it according to the comprehensive score of activity, contribution value and stability. The users with the highest activity, contribution value and stability comprehensive score in the same region are regarded as the first level nodes, the users with the highest activity, contribution value and stability comprehensive score in different regions of the same city are regarded as the second level nodes, and the users with the highest activity, contribution value and stability comprehensive score in all regions of other cities in the same province are regarded as the third level nodes. Finally, we design the cycle time to be T, and we will periodically adjust the node grading according to the location and comprehensive score ranking.

At present, there is no standard definition for the concept of "activeness" in the academic world. We reflect the activeness of users according to their average online time, number of messages sent, number of comments and number of likes. The user's activity is the product of the average online time in the cycle, the number of messages sent, the number of comments and the number of likes. The calculation formula is as follows (1):

$$activeness(u) = t * (k + m + n) \qquad (1)$$

Where, activeness (u) represents the activity of user u, t represents the average online duration of user u in the cycle, k represents the number of messages sent by user u, m represents the number of comments by user u, and n represents the number of likes by user u.

For the contribution value, the higher the cumulative storage cost, the higher the contribution value. The user's contribution value is the sum of the storage costs paid by users to help other users cache in the cycle. The calculation formula is shown in (2):

$$contribution(u) = \sum_{f \in v} w_f \qquad (2)$$

Where contribution (u) represents the contribution value of user u, Wf represents the storage cost of user cache file f, and V represents the collection of user u cache files.

For stability, we express that users who frequently go online and offline more often have low stability, and users who frequently go online and offline less often have high

Manuscript received Oct 05, 2022

Chenle Xiong School of Computer Science and Technology,Tiangong University,China.

stability[3]. The stability of a user is expressed as the reciprocal of the number of times a user goes online or offline. The calculation formula is shown in (3):

$$stability(u) = \frac{1}{freq} \quad (3)$$

Wherein, stability (u) represents the stable value of user u, and freq represents the online and offline times of user.

For the comprehensive score, we calculate the comprehensive score according to the scores of activity, contribution value and stability. The comprehensive score calculation formula is shown in (4):

score(u)=activeness(u)*contribuion(u)*stability(u) (4)

Where score (u) represents the comprehensive score of user u.

III. CACHE REPLACEMENT POLICY

Due to the limited cache space of user devices in decentralized social networks, when a user's cache space is saturated and cannot continue to cache content, users need to replace some content in the cache[4]. Therefore, this section will propose a cache replacement strategy based on content characteristics to solve this problem. Otherwise, when the user's cache space is saturated, new content cannot be cached, and the problem of outdated files occupying cache space for a long time cannot be solved, resulting in low user cache efficiency and slow network response.

In the cache replacement strategy (SNCR), we determine whether to replace the cache content by calculating the replacement coefficient based on the popularity of the content and the cache hit rate.

Content popularity refers to the popularity of a content. There are countless contents in social networks, but only a small part of popular content will be frequently accessed. When hot content is used to replace outdated content, it will greatly reduce the time delay for users to obtain content, improve the efficiency of users' retrieval, thus improving the user experience and reducing network pressure. We define the popularity formula as shown in (5):

$$P_T(CRY_i) = \alpha \times P_{T-2}(CRY_i) + \beta \times P_{T-1}(CRY_i) + \gamma \times RP_T(CRY_i) \quad (5)$$

Where P_ T (CRY_i) indicates the popularity of content i in cycle T, P_ (T-2) (CRY_i) indicates the popularity of content i in cycle T-2, RP_ T (CRY_i) indicates the relative popularity of content i in cycle T. α 、β 、γ Represent weight coefficient, meeting α +β+γ= 1 and 0 ⩽ α<β<γ ⩽ 1.RP_ The formula of T (CRY_i) is as follows (6):

$$RP_T(CRY_i) = \frac{n_T(CRY_i)}{\omega_0 \times n_{T-1}(CRY_i) + \omega_1 \times n_{T-2}(CRY_i)} \quad (6)$$

Where n_ T (CRY_i) indicates the number of times content i is requested in cycle T, n_ (T-1) (CRY_i) indicates the number of times content i is requested in cycle T-1, n_ (T-2) (CRY_i) indicates the number of times content i is requested in cycle T-2. ω_ 0、ω _ 1 represents the weight coefficient, and satisfies ω _ 0+ ω _ 1=1。

Cache hit ratio refers to the ratio of content hits and the sum of content hits and content misses. The formula is shown in (7):

$$r_T(CRY_i) = \frac{n_{Thit}(CRY_i)}{n_{Thit}(CRY_i) + n_{Tmiss}(CRY_i)} \quad (7)$$

Where r_ T (CRY_i) indicates the hit rate of content i in cycle T, n_ Thit (CRY_i) indicates the number of hits of content i in cycle T, n_ Tmiss (CRY_i) indicates the number of misses of content i in cycle T.

The replacement coefficient is obtained according to the content popularity and cache hit rate. The replacement coefficient formula is shown in (8):

$$Replace_T(CRY_i) = \mu_1 \times P_T(CRY_i) + \mu_2 \times r_T(CRY_i) \quad (8)$$

Where, Replace_ T (CRY_i) represents the replacement coefficient of content i in cycle T, μ _ 1、 μ _ 1 is the weight coefficient, meeting μ _ 1+ μ _ 2=1, and μ _ 1> μ _ 2。

The cache content replacement frequency refers to the frequency that a content is replaced by users in the network. The ratio of the number of content replacement times of the routing node to the total number of replacements in the cycle is recorded as the cache content replacement frequency. The calculation formula is shown in (9):

$$RF = \frac{Ts}{Ttotal} \quad (9)$$

Among them, RF represents the replacement frequency of cache contents, Ts represents the replacement times of routing nodes. When a content replacement occurs in the cache space of a node, the value of Ts will be increased by one.

In the cache replacement strategy designed in this paper, we get the replacement coefficient according to the popularity of the content and cache hit rate, and judge whether the content is worth caching by the replacement coefficient. The smaller the replacement factor is, the less popular the content is in the network, and the content will be replaced first[5].

IV. CONCLUSION

The cache content of decentralized social networks requires users in the network to dedicate their own device space to cache this part of content, but users are not great and everyone is selfish, so we need to design incentive mechanisms to encourage users to cache, so as to ensure smooth operation of social networks. In the follow-up work, we will propose an effective incentive mechanism to encourage users in the network to cache, so as to ensure the smooth network.

REFERENCES

[1] Synereo Partners with China's Qiyi Culture[J]. Manufacturing Close - Up, 2018.

[2] Ren Jifan et al. Relationship between consumer participation behaviors and consumer stickiness on mobile short video social platform under the development of ICT: based on value co-creation theory perspective[J]. Information Technology for Development, 2021, 27(4) : 697-717.

[3] Hanami YOKOI,Takuji TACHIBANA. Autonomous Relay Device Placement Algorithm for Avoiding Cascading Failure in D2D-Based Social Networking Service:Special Section on the Architectures,

Protocols, and Applications for the Future Internet[J]. IEICE Transactions on Information and Systems,2021,E104.D(5).
[4] Mega Giuliano and Montresor Alberto and Picco Gian Pietro. Social Overlays Meet the Cloud: A Hybrid Architecture for Profile Dissemination in Decentralized Social Networks[J]. IEEE Transactions on Network Science and Engineering, 2019, 6(4) : 613-627.
[5] Peer-to-Peer Networking; Data on Peer-to-Peer Networking Reported by Researchers at Autonomous University (A performance comparison of Chord and Kademlia DHTs in high churn scenarios)[J]. Network Weekly News, 2015.