

성능 및 유사도 정보를 이용한 슈퍼 피어 선별 기법

Intelligent Capacity and Similarity based Super-peer Selection in P2P Network

*민수홍, **조동섭
Suhong Min, Dongsub Cho

Abstract - The peer-to-peer (P2P) systems have grown significantly over last few years due to their high potential of sharing various resources. Super-peer based P2P systems have been found very effective by dividing the peers into two layers, SP (Super-Peer) and OP (Ordinary-Peer). In this paper, we present ISP2P (Intelligent Super-peer based P2P system), which allows us to choose the best SP. Through analyzing capacity and similarity between SP and OP, we can help OPs to select the most appropriate SP respectively. Proposed system can improve the performance of the average response time by superior SP, reduce the bandwidth cost by small path length due to content similarity and solve frequent SP replacement problem by considering similarity of user behavior.

Key Words :Peer-to-Peer (P2P), Super-peer, Capacity, Similarity

1. 장 서 론

최근의 P2P 시스템들은 초기 순수 P2P 모델에서 슈퍼 피어 기반의 계층형 모델로 변화되고 있다. 순수 P2P 모델은 중앙의 서버 없이 피어와 피어가 자율적으로 원하는 자원을 찾는 방식이었다. 그러나 이러한 모델은 피어가 원하는 자원을 찾기 위해서 네트워크에 많은 메시지를 발생시켜야 하며, 또 원하는 자원의 위치를 정확히 찾는 데 어려움이 따른다. 이러한 문제를 해결하기 위해서 최근 슈퍼 피어를 기반으로 하는 계층형 P2P 모델에 대한 연구가 활발히 진행되고 있다. 슈퍼 피어 기반의 P2P 모델에서는 네트워크의 계층을 성능이 우수한 슈퍼 피어와 일반 피어로 분류한다 [1]. 슈퍼 피어는 일반 피어의 모든 쿼리를 전담하여 처리함으로써 성능이 낮은 피어의 네트워크에 참여를 도울 수 있다. 기존의 중앙 집중형 방식과 달리 슈퍼 피어 기반의 시스템은 여러 슈퍼 피어들이 존재하여 일반 피어들을 슈퍼 피어를 중심으로 분산되도록 한다. 슈퍼 피어 기반의 P2P 시스템의 특징은 일반 피어가 자원 공유를 위해 반드시 하나의 슈퍼 피어를 선택해야 하며, 선택된 슈퍼 피어를 통해 네트워크에 참여할 수 있다. 따라서 슈퍼 피어는 일반 피어의 쿼리를 잘 처리할 수 있어야 하며, 일반 피어가 요구하는 자원을 잘 검색할 수 있어야 한다. 기존의 연구들은 슈퍼 피어를 선택하는 문제에 대해 간과되고 있다. 본 논문에서는 슈퍼 피어를 선별하기 위해 일반 피어 대 슈퍼 피어의 물리적인 성능과 콘텐츠 및 행위에 대한 유사도를 측정한다.

제안한 시스템은 다음과 같은 기여도를 갖는다.

- 성능이 가장 우수한 슈퍼 피어를 선택함으로써 많은 쿼리들을 효율적으로 처리할 수 있어 쿼리 처리의 응답 시간을 향상시킬 수 있다.
- 유사한 콘텐츠를 처리하는 슈퍼 피어를 제공함으로써 짧은 경로 안에 자원을 찾을 수 있도록 함으로써 대역폭 낭비를 향상시킬 수 있다.
- 유사한 사용 시간대에 사용하는 슈퍼 피어의 선택으로 인해 빈번한 슈퍼 피어 교체 문제를 해결할 수 있다.

2. 장 관련 연구

초기 그누텔라 0.4 버전에서는 순수 P2P 구조로써, 피어가 다른 피어와의 연결 정보를 얻기 위해서 Ping 메시지를 현재 연결된 모든 피어에게 전송한다. 수신한 피어들은 응답 값으로 자신의 연결 정보와 자원을 pong 메시지로 전송하며, 또한 수신한 ping 메시지를 자신과 연결된 다른 피어에게 브로드 캐스팅 한다. 따라서 각각의 들이 메시지를 핸들링하고, 라우팅 해야 하므로 피어의 수가 증가함에 따라 네트워크 내의 트래픽 또한 크게 증가된다. 그누텔라는 이러한 문제점을 해결하기 위해 버전 0.6에서는 슈퍼 피어 개념인 ultrapeer를 사용한다. ultrapeer는 proxy로서 동작하는데, 피어들의 쿼리를 QRP(Query Routing Protocol)을 사용해서 대신 전달하는 역할을 한다. 그누텔라에서는 ultrapeer와 일반 피어를 분류할 때 피어의 CPU 성능과 네트워크 성능에 제약 사항을 두고 이를 만족할 경우, ultrapeer로 분류한다[2]. 이처럼 그누텔라에서는 슈퍼 피어들이 자체의 성능으로 분류되며, 일반 피어는 슈퍼 피어를 선택할 때 ping 메시지를 전송해 응답하는 슈퍼 피어 중 랜덤한 방식으로 슈퍼 피어를 선별 한다 [2].

* 민수홍 : 이화여자대학교 컴퓨터학과 박사과정

** 조동섭: 이화여자대학교 컴퓨터학과 정교수

3. 장 시스템 설계

3.1 절 시스템 특징

제안한 시스템의 특징은 다음과 같다. 첫째, 일반 피어가 최적의 슈퍼 피어를 선택할 수 있도록 일반 피어 대 슈퍼 피어의 성능을 분석한다. 둘째, 슈퍼 피어와 일반 피어의 공통된 특징, 즉, 콘텐츠 유사도와 사용 행위 유사도를 분석한다. 셋째, 최적의 슈퍼 피어를 선택하기 위해서 성능 값과 유사도 값을 가중치에 따라 일반 피어에 대한 각각의 슈퍼 피어의 점수를 계산한다. 전체 시스템의 구조는 그림 1과 같다.

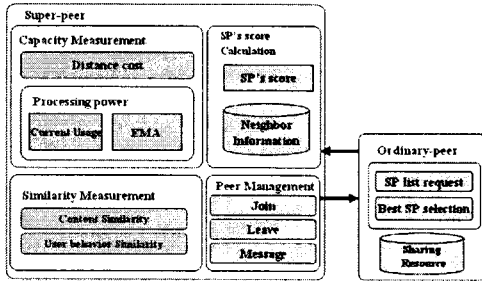


그림 1 시스템 구조

3.2 절 성능 측정

제안된 시스템은 베스트 슈퍼 피어를 선택하기 위해서 성능 요소를 고려할 때 슈퍼 피어의 성능 그 자체를 다루는 것이 아니라 일반 피어 대 슈퍼 피어 사이의 성능을 고려하였다. 즉, 일반 피어와 슈퍼 피어의 거리와 일반 피어가 슈퍼 피어를 선택했을 때의 예상되는 슈퍼 피어의 로드값을 고려하였다. 이를 위해서 먼저 일반 피어와 슈퍼 피어 사이의 거리 비용을 측정하고, CPU 처리 능력을 측정한다 [3].

3.2.1 절 거리 비용

링크 사이의 거리 비용은 전송 지연과 전파 지연, 그리고 홉을 이용해 계산한다. 현재 기존의 시스템들은 홉에 의존한다. 그러나 홉의 수는 물리적인 거리가 가까울지라도 홉 수는 더 클 수 있는 점을 고려해 제안한 논문에서는 라우터당 전송 지연, 전파 지연 그리고 홉을 계산한다.

$$T_d = \text{Packetsize} / \text{Rate} \quad (1)$$

$$P_d = \text{PacketLength} / 2 * 10^8 \quad (2)$$

$$D_v = \sum_{i=0}^{\max \text{ hop}} (T_d + P_d) \quad (3)$$

총 거리비용 D_v 는 전송 지연 T_d 과 전파 지연 P_d 에 대해 각 홉 수만큼 계산해서 구한다 [3,4].

3.2.2 절 CPU 성능

CPU 로드는 사용자의 네트워크 상태, 이용 행위 또는 갑자기 발생한 트래픽에 따라 동적으로 변한다. 따라서 일반 피어가 슈퍼 피어에게 CPU 로드를 요구할 때 현재의 CPU

로드는 슈퍼 피어의 CPU 로드를 정확히 반영한다고 보기 어렵다. 이를 위해 본 논문에서는 슈퍼 피어의 성능을 구하기 위해 EMA(Exponential Moving Average)를 수정한다. 제안한 방법은 주기를 3단계로 나누어서 CPU 로드를 계산하였다. 먼저 가장 짧은 주기로 현재의 CPU 로드를 측정한다. 마지막으로 가장 긴 주기에 따라 EMA를 구하며, 이에 따라 계산된 결과가 CPU 성능 비용이 된다.

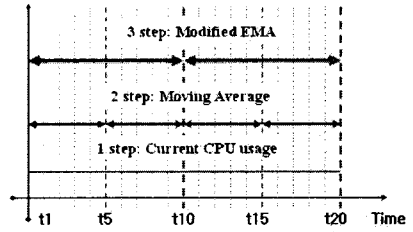


그림 2 단계별 CPU 로드 측정

$$C_t = \{c_1, c_2, c_3, \dots, c_N\} \quad (1)$$

$$MA_t = \frac{\sum_{i=0}^{n-1} C_{t-i}}{n} \quad (2)$$

$$C_\mu = \alpha \cdot MA_t + (1 - \alpha) \cdot EMA_{t-1} \quad (3)$$

3.3 절 유사도 측정

일반 피어는 콘텐츠와 사용 행위에 대해 공통된 특징을 가지고 있는 슈퍼 피어를 선택하도록 한다. 콘텐츠 기반 유사도는 일반 피어가 슈퍼 피어를 선택했을 때 해당 슈퍼 피어가 일반 피어에게 가장 관련된 콘텐츠를 제공할 수 있는지의 여부를 살펴본다. 선택한 슈퍼 피어가 유사한 콘텐츠의 제공이 가능한 경우, 일반 피어는 보다 효율적으로 적은 경로를 거쳐 자원을 획득할 수 있으며 대역폭 낭비도 줄일 수 있다. 또한 일반 피어는 가장 유사한 행위 즉 유사한 시간대, 유사한 이용 시간을 보이는 슈퍼 피어를 선택 함으로써 이를 통해 잦은 슈퍼 피어 교체 문제를 해결할 수 있다.

3.3.1 절 콘텐츠의 유사도 측정

제안한 시스템에서는 콘텐츠에 대한 유사도를 측정하기 위해서 ISM 알고리즘[5,6]을 이용하였다. 이 알고리즘은 Nearest Neighbor 분류 알고리즘과 코사인 유사도를 사용한다. 각각의 콘텐츠의 유사도에 대한 점수를 다음과 같이 계산한다.

$$Psimp_k(P_i, q) = \sum Qsim(q_j, q)^\alpha \quad (1)$$

일반 피어가 슈퍼 피어에게 자원을 검색할 때 슈퍼 피어는 해당 쿼리에 대해 콘텐츠 유사도를 측정해 슈퍼 피어가 가장 관련된 정보를 가지고 있는 일반 피어 또는 슈퍼 피어의 정

보를 제공한다.

3.3.1 절 이용 행위에 대한 유사도 측정

일반 피어와 슈퍼 피어의 P2P 서비스를 이용하는 행위에 대한 유사도 정보를 알기 위해서 본 논문에서는 사용 시간대와 이용 시간에 따라 슈퍼 피어와 일반 피어의 유사도 값을 계산한다. 슈퍼 피어 기반의 P2P 시스템의 경우, 일반 피어는 슈퍼 피어의 네트워크 존재 여부에 따라 영향을 받는다. 슈퍼 피어가 일반 피어와 다른 시간대에 존재하거나 또는 이용 시간이 일반피어에 비해 현저하게 작을 경우, 슈퍼 피어의 잦은 네트워크 탈퇴로 인해 일반 피어는 서비스 중간에 슈퍼 피어를 교체해야 한다. 본 논문에서는 일반 피어가 슈퍼 피어를 선택할 때 P2P를 이용하는 시간대와 같은 시간대에 대해서 사용 시간이 유사한 슈퍼 피어를 선택하도록 한다. 방법은 다음과 같다. 먼저, 슈퍼 피어와 일반 피어 각각 주로 사용하는 사용 시간대를 알아보고, 둘째, 각각의 시간대에 따라 평균 이용 시간을 계산한다. 이를 위해서 우리는 먼저, 시간은 GMT 를 기준으로 하여 시간대를 8 zone 으로 분류하였다.

| 타임존 | 시간 (A.M) | 타임존 | 시간 (P.M) |
|-----|----------|-----|----------|
| 1 | 0 ~ 3 | 5 | 12 ~ 3 |
| 2 | 3 ~ 6 | 6 | 3 ~ 6 |
| 3 | 6 ~ 9 | 7 | 6 ~ 9 |
| 4 | 9 ~ 12 | 8 | 9 ~ 12 |

<표 1> GMT 기반의TIME ZONE

일반 피어와 슈퍼 피어는 각각 이용하는 평균 시간대를 계산한 다음, 이를 통계적으로 가장 많이 사용하는 순으로 정렬한다. 일반 피어는 각각의 타임존에 대해 가중치를 적용해 슈퍼 피어의 타임존 값과 비교한다. 식은 다음과 같다.

$$Tsim_N = \frac{\sum_{i=0}^N \alpha(n-i)(p_i \cap q^\alpha)}{\sum_{i=0}^n (n-i)} \quad (1)$$

3.4 절 슈퍼 피어 선별

슈퍼 피어를 선별하기 위해서 일반 피어는 일반 피어 대 슈퍼 피어의 최대 성능, 최대 유사도를 계산한다.

- 최대 성능: 슈퍼 피어는 일반 피어와의 거리 비용과 CPU 성능과 같은 물리적인 성능을 계산한다. 전체 슈퍼 피어와의 거리 비용이 D라고 할 때 하나의 슈퍼 피어와의 거리 비용은 d이며, $d \in D$ 라고 표현한다. CPU 성능은 전체 P라고 할 때 $p \in P$ 에 속한다. 최대 성능은 $MC: d_i \times p_i \mapsto [0,1]$ 로 나타낼 수 있으며, 만약 슈퍼 피어의 거리 비용 d_i 가 0.8이고, CPU 비용이 0.5 라고 가정하면, 최대 성능은 0.4이다.

- 최대 유사도: 슈퍼 피어는 콘텐츠 유사도와 사용자 행위에 대해 최대 유사도를 계산한다. 각각의 콘텐츠 유사도가 c 이면, c는 전체 유사도 집합 C에 속하며 사용자 행위를 u라고 할 때 $u \in U$ 로 표현된다. 최대 유사도는

$MS: c_i \times u_i \mapsto [0,1]$ 로 나타낸다. 만약 유사도 값이 0이면 일반 피어 대 슈퍼 피어 사이는 유사성이 전혀 없으며, 1이면 완전히 두 피어가 일치한다고 본다.

일반 피어는 슈퍼 피어를 선별하기 위해서 다음과 같이 계산한다. $SP_i: MS_i \times MC_i \mapsto [0,1]$ 이며, 만약 각각의 슈퍼 피어 SP1이 0.5, SP2가 0.8이면, 일반 피어는 베스트 슈퍼 피어로서 SP2를 선택한다.

4. 장 결론

본 논문에서는 일반 피어 대 슈퍼 피어의 성능 및 유사도를 분석해서 최적의 슈퍼 피어를 선택할 수 있도록 제안하였다. 기존의 슈퍼 피어 기반의 P2P 시스템에서는 일반 피어가 슈퍼 피어를 랜덤하게 선택하거나 슈퍼 피어의 선택의 문제를 간과하고 있다. 그러나 전체 P2P 시스템에서 일반 피어의 모든 쿼리를 대신 처리해야 할 슈퍼 피어의 역할은 매우 중요하며, 이를 위해서 일반 피어가 최적의 슈퍼 피어를 선택할 수 있어야 한다. 본 논문에서는 크게 성능과 유사도를 기반으로 슈퍼 피어를 선별하도록 한다. 제안한 시스템은 성능의 고려를 통해 쿼리 처리를 위한 평균 응답 시간을 항상 시킬 수 있으며, 콘텐츠 유사도를 통해 대역폭 문제를 해소하며, 유사한 사용 시간을 분석해서 잦은 슈퍼 피어의 교체 문제를 해결할 수 있다.

참 고 문 헌

- [1] B. Yang, H. Garcia-Molina, "Designing a super-peer network", IEEE International Conference on Data Engineering, Bangalore, India, March 2003.
- [2] Gnutella protocol spec. v.0.6
http://rfc-gnutella.sourceforge.net/src/rfc-0_6-draft.html
- [3] Suhong. Min, D. Cho, "An Intelligent Performance based Hybrid P2P System", Journal of Korea Electrical Engineering and Technology", 5(2). 2006. 2.
- [4] J.F.Kurose, K.W.Ross, "Computer Networking: A Top-Down Approach Featuring the Internet", Addison Wesley, 2002.
- [5] V. Kalogeraki, D. Gnuopulos and D. Zeinalipour-Yazti, "A Local Search Mechanism for Peer-to-Peer Networks", Proceedings of CIKM'02, McLean VA, USA, 2002.
- [6] D. Tsoumakos, N. Roussopoulos, "A comparison of peer-to-peer search methods", In proceedings of the sixth International Workshop on the Web and Database, SanDiago, CA, 2003.