

## 빅데이터를 위한 라우팅 메시지 재전송 기법 연구

이병준<sup>○</sup>, 윤희용<sup>\*</sup>

<sup>○\*</sup>성균관대학교 정보통신대학

e-mail: {bjlee<sup>○</sup>, youn<sup>\*</sup>}@ece.skku.ac.kr

### A Study on Routing Message Retransmission Scheme for Big data

Byung-Jun Lee<sup>○</sup>, Hee-Yong Youn<sup>\*</sup>

<sup>○\*</sup>College of Information and Communication Engineering, Sungkyunkwan University

#### ● 요약 ●

최근 소셜 네트워크 서비스로 대표되는 정보유통 매체의 급격한 발전으로 인해 데이터 빅뱅(Data Big Bang)이라 할 수 있는 데이터의 폭발적인 증가 현상에 따라 빅데이터에 대한 관심이 급격히 증대되고 있다. 빅데이터 관련 기술들은 기본적으로 대용량 데이터를 하나의 노드로 관리하는 것이 아닌 여러 노드를 연결하기 때문에 효율적인 데이터 관리를 위해서 노드 간 연결을 담당하는 라우팅 알고리즘의 중요성 역시 대두되고 있다. 본 논문에서 대용량 데이터를 위한 효율적 라우팅 알고리즘을 위해 새로운 라우팅 메시지 재전송 기법을 위한 혼잡 확률 연산 알고리즘을 제안한다.

키워드: 빅데이터(Big data), 라우팅(Routing), 재전송(Retransmission)

#### I. 서론

빅데이터란 단순히 대용량의 데이터만을 의미하는 것이 아니라 대용량의 데이터를 활용, 분석하여 가치 있는 정보를 추출하고 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 정보화 기술을 말한다.

이런 빅데이터를 처리하기 위한 관련 기술들은 주로 방대한 데이터를 처리하기 위하여 여러 노드를 이용한 분산처리 시스템을 사용한다. 이렇게 여러 노드를 효율적으로 활용하기 위해서 Directed Diffusion 데이터 중심 라우팅 기법이나 최소 싱크간 최소 비용 측정치 기반 포워딩 라우팅 기법을 사용하는 경우가 있다. 그러나 이런 기법들은 라우팅을 위해 메시지 전송 시 다중 경로 중 플러딩을 막기 위해 전송품질이 좋은 노드들만을 이용하거나 메시지 전송에 실패할 시 일정 시간을 경과한 후 다시 메시지를 재전송하므로 효율적인 라우팅 서비스를 제공하지 못할 경우가 있다. 본 논문에서는 메시지 전송 실패의 원인을 분석하고 재전송하기 위해 대기시간을 유동적으로 조절하는 기법을 제안한다.

#### II. 관련 연구

##### 1. 빅데이터 관련 기술

대표적인 빅데이터 관련 기술로는 아파치 하둡 에코시스템(Hadoop Eco System)을 들 수 있다[1]. 하둡은 대량의 자료를

처리할 수 있는 큰 컴퓨터 클러스터에서 동작하는 분산 응용 프로그램을 지원하는 시스템이다. 이 시스템은 분산파일 시스템인 HDFS(Hadoop Distributed File System)과 맵리듀스(MapReduce)라는 이름의 분산처리시스템으로 구성되어 있으며 최근 하둡2.0을 발표하였다[2].

##### 2. 기존 라우팅 알고리즘

빅데이터 처리 시스템에 사용 가능한 기존 라우팅 기법으로는 Directed Diffusion 데이터 중심 라우팅 기법[3]과 최소 싱크간 최소 비용 측정치 기반 포워딩 기법[4]이 있다. Directed Diffusion은 발생 가능한 플러딩을 막기 위해 경로들 중 전송품질이 좋은 몇 개의 경로들이 강화되어, 강화된 경로를 통해 데이터 전송이 이루어진다.

최소 싱크간 최소 비용 측정치 기반 포워딩 기법은 노드에서 여러 번의 갱신이 일어나거나 여러 번의 전송을 수행하고, 싱크에서 멀리 떨어진 노드들의 더 많은 갱신이 일어날 수 있는 점을 해결하기 위하여, 현재 비용 측정치를 갱신할 때 백 오프 기법을 적용하여, 갱신 시간으로부터 광고 메시지를 수신한 링크 비용의 정수배의 시간이 경과한 후에 광고 메시지를 전송 하도록 한다.

#### III. 본론

Node Condition(NC)은 노드 간 상태를 가져오기 위해 사용된다.

Average Message Error Rate (*aMER*)은 NC에 수집된 오류율들의 평균을 낸 것이고 Previous *aMER*(*pMER*)은 현재의 *aMER*값과 비교하기 위한 변수이다. Bad Threshold(*BT*)와 Good Threshold(*GT*)는 노드 간 상태가 좋고 나쁨을 구분하기 위한 메시지 오류율로 각각 18%와 5%로 정해져 있다. 이는 노드 간 혼잡인 경우에 보다 능동적으로 대처하기 위함이다. Query Number(*QN*)은 노드 간 상태를 가져올 때 질의의 수를 나타내며 Cong은 질의를 통한 노드 간 상태를 이용하여 계산된 메시지 손실이 혼잡으로 인해 발생될 확률이다. 이 확률은 일정시간 간격으로 계산되며 혼잡확률을 수시로 변경한다. 그림1에서 노드 간 상태 연산 기법의 순서도를 보여주며 이렇게 연산된 노드 간 메시지 손실 원인이 혼잡일 확률은 Cong로 표현한다.

#### IV. 결론

본 논문에서는 빅데이터 처리를 위한 시스템 내에서 효율적으로 라우팅하기 위한 방안으로 노드 간 혼잡 확률 연산 기법을 제안했다. 제안된 기법은 여러 노드를 메시지 전송 실패의 원인을 분석하여 보다 효과적인 라우팅을 지원할 수 있다. 향후 혼잡확률 연산 기법을 이용하여 라우팅 메시지 재전송 타이머 기법에 대한 연구가 필요하다.

estimation scheme

#### ACKNOWLEDGMENT

본 연구는 한국산학연합회(C0017380), BK21+사업, 한국연구재단 기초연구사업 (2013R1A1A2040257), (2013R1A1A2060398), 미래부가 지원 한 2013년 정보통신-방송(ICT) 연구개발 사업 (1391105003)의 지원을 받아 수행되었음.

#### 참고문헌

- [1] HADOOP Project, <http://hadoop.apache.org>
- [2] HADOOP 2.0 Project, <http://hadoop.apache.org/docs/stable/>
- [3] Casey, K., Neelisetti, R., Lim, A., "RTDD: A Real-Time Communication Protocol for Directed Diffusion," Wireless Communications and Networking Conference, pp.2852-2857, 2008.
- [4] Henderson, W.D., Tron, S., "Verification of The Minimum Cost Forwarding Protocol for Wireless Sensor Networks," Emerging Technologies and Factory Automation, pp.194-201, 2006.

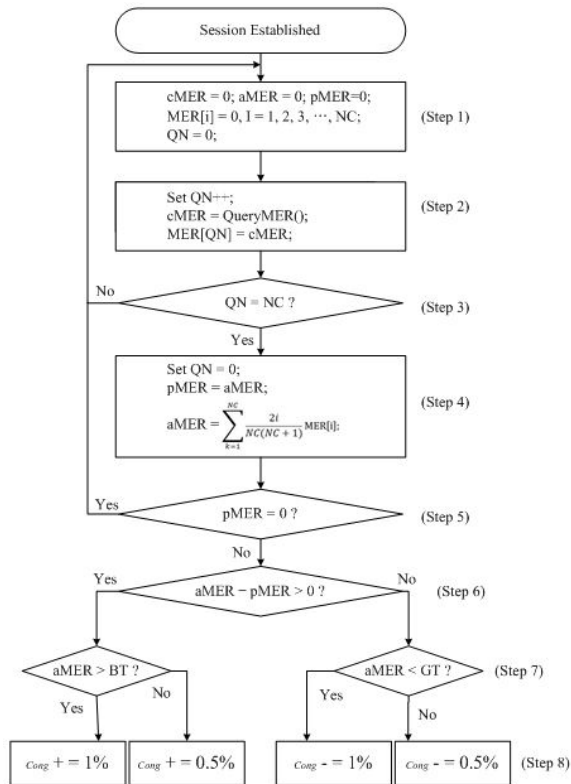


그림 1. 혼잡 확률 연산 순서도

Fig. 1. The flow chart of congestion probability