Load Adaptive and Fault Tolerant Distributed Stream Processing System for Explosive Stream Data

Myungcheol Lee*, Miyoung Lee*, Sung Jin Hur*, Ikkyun Kim**

*Big Data SW Research Department, ETRI(Electronics and Telecommunications Research Institute),
218 Gajeong-ro, Yuseong-gu, Daejeon, 305-700, Republic of Korea

**Cyber Security System Research Department, ETRI(Electronics and Telecommunications Research Institute),
218 Gajeong-ro, Yuseong-gu, Daejeon, 305-700, Republic of Korea

{mclee, mylee, sjheo, ikkim21}@etri.re.kr

Abstract—As smart devices such as sensors, smartphones, and CCTVs are becoming extensively utilized recently, stream data from those smart devices are consistently generated explosively. There are also increasing cases that we notice security attacks after already important assets are damaged by cyber-targeted attacks such as APT attacks due to the lack of real-time security log processing capability. Accordingly, the demand to process and analyse the exploding stream data in real-time and in advance is consistently increasing in many application domains. However, existing distributed stream processing systems like Storm and S4 are not well adaptive when there are drastic increase of input stream data. In this paper, we propose a distributed stream processing system which supports several load adaptation techniques utilizable for various circumstances of explosive data stream, and also supports fault tolerance mechanisms to fail over in several failure situations.

Keyword—Big Data, Data Explosion, Distributed Stream Processing, Fault Tolerance, Load Adaptation, Load Shedding, Task Scheduling
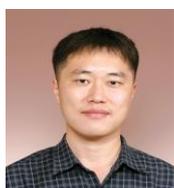
Myungcheol Lee (M'2015) received his Bachelor's Degree in computer engineering and his Master's Degree in computer engineering from Chungnam National University, Daejeon, Korea in 1999 and 2001, respectively. He became a Member (M) of IEEE in 2015. He is now a senior researcher at ETRI since 2001. His research interest includes Big Data management, processing and analytics, database, cloud computing, and distributed computing.

Miyoung Lee received her Bachelor's Degree in food and nutrition, and Master's Degree in computer science and statistics from Seoul National University, Seoul, Korea in 1981 and 1983, respectively. She received her Doctor's Degree in computer engineering from Chungnam National University, Daejeon, Korea in 2005. She is now a principal researcher at ETRI since 1988. Her research interest includes big data management and processing, database, and distributed computing.

Sung Jin Hur received his Bachelor's Degree in electronics, Master's Degree and Doctor's Degree in computer engineering from Kyungpook National University, Daegu, Korea in 1990, 1992 and 1999, respectively. He was a professor at Changsin University from 1999 to 2001, and is now a principal researcher at ETRI since 2001, and is leading Data Management Research Section. His research interest includes database, stream data processing, cloud computing.

Ikkyun Kim received his Bachelor's Degree, Master's Degree and Doctor's Degree in computer engineering from Kyungpook National University, Daegu, Korea in 1994, 1996, and 2009, respectively. He was a visiting researcher at Purdue University from 2004 to 2005. He is now a principal researcher at ETRI since 1996, and is leading Network Security Research Section. His research interest includes network security, computer network, cloud security, big data analytics.