DuDE: A Prototype for a P2P-based Distributed Computing System

Jan Skodzik, Peter Danielis, Vlado Altmann,
Jens Rohrbeck, Dirk Timmermann
University of Rostock
Institute of Applied Microelectronics and Computer Engineering
18051 Rostock, Germany
Tel./Fax: +49 (381) 498-7272 / -1187251
Email: jan.skodzik@uni-rostock.de

Thomas Bahls, Daniel Duchow
Nokia Siemens Networks GmbH & Co. KG
Broadband Access Division
17489 Greifswald, Germany
Tel: +49 (89) 5159-22771 / -18237
Email: {thomas.bahls;daniel.duchow}@nsn.com

Abstract—The P2P-based system for the distributed calculation of statistics called DuDE is presented. It is implemented for PCs with Windows OS and is highly configurable regarding the functional spectrum and many parameters concerning individual functionalities. A GUI has been developed for testing and demonstrating DuDE. We show the DuDE functionality in a demonstration scenario and discuss several alternative constellations during the demonstration.

I. INTRODUCTION

To analyze bottlenecks, detect attacks on parts of the Internet Service Providers (ISPs) networks, or to determine the state of components, ISPs use log data traced by their hardware. Furthermore, generated statistics allow the ISP to raise the quality of service of their networks. The creation of statistics is performed by central computation systems. These systems are not scalable and will be overwhelmed with the ever increasing log data volumes [1]. In addition, solely a standard set of statistics can be generated for a single access node (AN). The computation of long term statistics (LTS) is not supported at all. To remedy these deficiencies, DuDE—a Distributed Computing System using a Decentralized P2P Environment— has been developed. ANs of an ISP's access network are connected by a self-organizing distributed hash table (DHT)-based P2P network superseding an additional centralized system for statistics calculation. These ANs have a certain unused available storage and computing capacity. These resources may be used at no extra costs, assuming some idle time or storage capacity being left in an average AN. Dynamic load balancing ensures that only nodes with enough resources are used to calculate statistics depending on their currently available resources. Therefore, the overloading of single nodes is avoided. DuDE scales with increasing log data volumes in contrast to a centralized solution. In addition, DuDE guarantees high data availability by using P2P technology with its high resilience [2] in combination with sophisticated redundant data storage. DuDE supports the calculation of complex and LTS—both for a single AN and for all nodes.

The system is implemented as software prototype on PCs with Windows OS. Thereby, the PC emulates an AN with DuDE functionality. To meet different demands, DuDE is
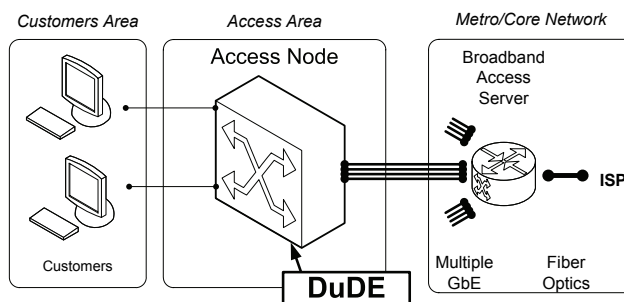

Fig. 1. Access network with DuDE node.

highly configurable at compilation time, e.g., regarding the actual functional spectrum and many parameters concerning the individual functionalities. Furthermore, a graphical user interface (GUI) has been developed to test and demonstrate the DuDE prototype.

Section II presents the DuDE prototype. The main characteristics of DuDE are explained. In Section III, the demonstration and test tool is introduced. Section IV briefly sketches a demonstration scenario of the DuDE working process before the paper concludes in Section V.

II. DuDE – THE P2P-BASED DISTRIBUTED COMPUTING SYSTEM

A typical access network depicted in Figure 1 comprise so called ANs like IP DSL Access Multiplexers (IP DSLAMs). DuDE is located on these ANs. Kad is the DHT-based P2P network, which has been chosen for the implementation of DuDE [3]. Thereby, all ANs are organized into a Kad-based DHT.

ANs participating in the statistics calculation can become job scheduler and/or task watcher depending on their available resources. A job scheduler is a very powerful node responsible for distributing parts of the statistics calculation (tasks) to ANs with sufficient available resources. For each new request from the user, the job scheduler is determined again to be able to adapt to changing network conditions. Each task watcher calculates its statistics part and returns it to the job scheduler
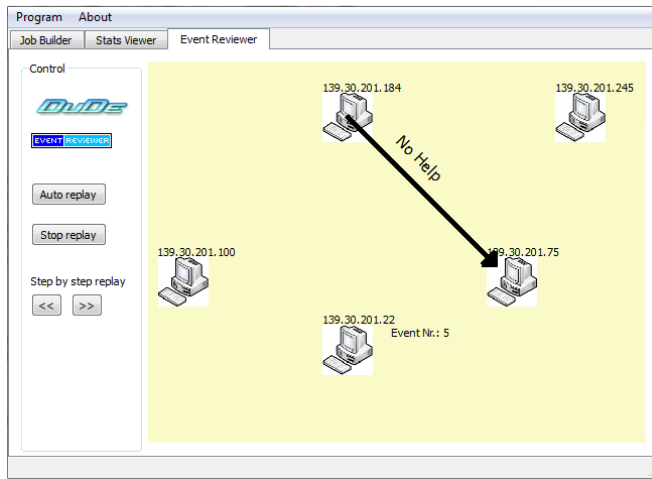
Fig. 2.   DuDE demonstration and test tool.

Fig. 3.   Displaying generated statistics with gnuplot.

after completion. An ISP administrator can request calculated statistics from any AN and does not need to be connected to the job scheduler. To keep communication overhead at a minimum, the usage of migration is renounced.

III. THE DEMONSTRATION AND TEST TOOL

To demonstrate and test DuDE's functionality, a graphical user interface (GUI) has been developed. The screenshot in Figure 2 shows the GUI of the tool's latest version. It is developed in C++ using the open source version of Nokia's Qt [4]. The first tab "Job Builder" of the GUI allows for the configuration of statistics to be generated and starts the calculation (job). During the job, selected statistics are computed. The second tab "Stats Viewer" is dedicated to show generated statistics. It is possible to choose actual values of nodes or a whole functional profile of parameters. The GUI dynamically creates gnuplot scripts depending on the requested statistics. Figure 3 shows calculated statistics created with a dynamically generated gnuplot script. The actual CPU usage of all nodes in the P2P network is depicted. The x axis shows the node ID of the node, which is a unique identifier for each node. All files are assigned to a node and contain the node ID in their filenames. Data is displayed in an extra gnuplot window. The third tab "Event Reviewer" depicts the communication sequence of involved nodes. An extra log file traces the communication between the nodes. It is possible to review the whole communication from the perspective of one node. The GUI is connected via local host to the DuDE instance and communicates via UDP packages.

IV. DuDE DEMONSTRATION SCENARIO

DuDE has been developed with Visual Studio 2008 and runs on Windows OS. The complex environment of an AN cannot be rebuilt. Instead, a software prototype written in C++ emulates the AN with DuDE functionality. The P2P functionality is fully integrated into the prototype.

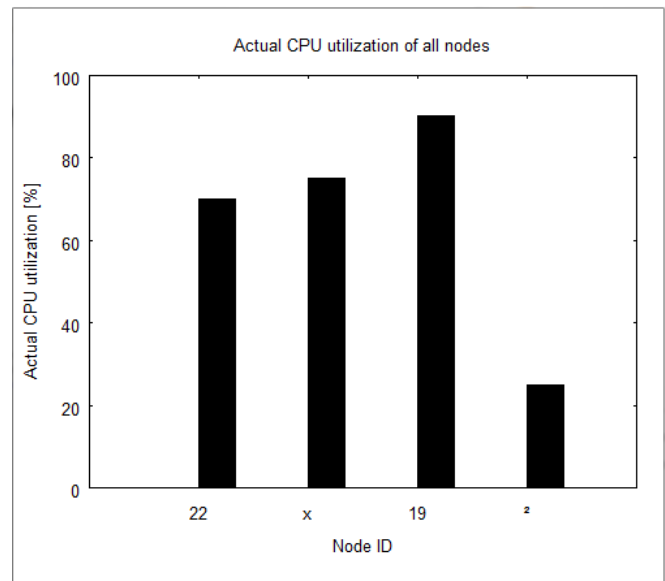**DuDE Demonstration Scenario:** Three PCs are connected via the Kad protocol. The free processor and memory resources of the DuDE nodes are hard coded. This allows the user to assign job scheduler and task watcher indirectly. The first DuDE node is connected to the terminal, which means the GUI is running on it. The second DuDE node has most free resources and becomes the job scheduler. The third node computes the task. Node one has least resources and is not involved in the statistic computing. Generated statistics are sent back to the first DuDE node and can be displayed via gnuplot.

V. CONCLUSION

We presented DuDE as a working P2P-based distributing computing system to create statistics. We achieved a fully scalable system by utilizing P2P technology. The DuDE GUI supports the creation of jobs and interpretation of generated statistics. As third aspect, the GUI allows the user a graphical communication replay. The working process is not just limited to generate statistics. DuDE can meet a various number of requirements like the fast integration of new statistic types or solving different computationally expensive tasks.

REFERENCES

[1] Traffic Statistics by De-CIX. [Online]. Available: https://www.de-cix.net/content/network/Traffic-Statistics.html
[2] R. Steinmetz and K. Wehrle, *Peer-to-Peer Systems and Applications*, ser. Lecture Notes in Computer Science.   Springer-Verlag Berlin Heidelberg, 2005.
[3] P. Maymounkov and D. Mazieres, "Kademlia: A peer-to-peer information system based on the xor metric."   IPTPS, 2002.
[4] Qt by Nokia. [Online]. Available: http://qt.nokia.com/