

Self-learning classifier for internet traffic

*Original*

Self-learning classifier for internet traffic / Ram, Keralapura; Mellia, Marco; Grimaudo, Luigi. - (2014).

*Availability:*

This version is available at: 11583/2540288 since:

*Publisher:*

*Published*

DOI:

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

## SELF-LEARNING CLASSIFIER FOR INTERNET TRAFFIC

Ram Keralapura, San Jose, CA (US); Marco Mellia, Turin (IT); and Luigi Grimaudo, Turin (IT)

Assigned to Narus, Inc., Sunnyvale, CA (US)

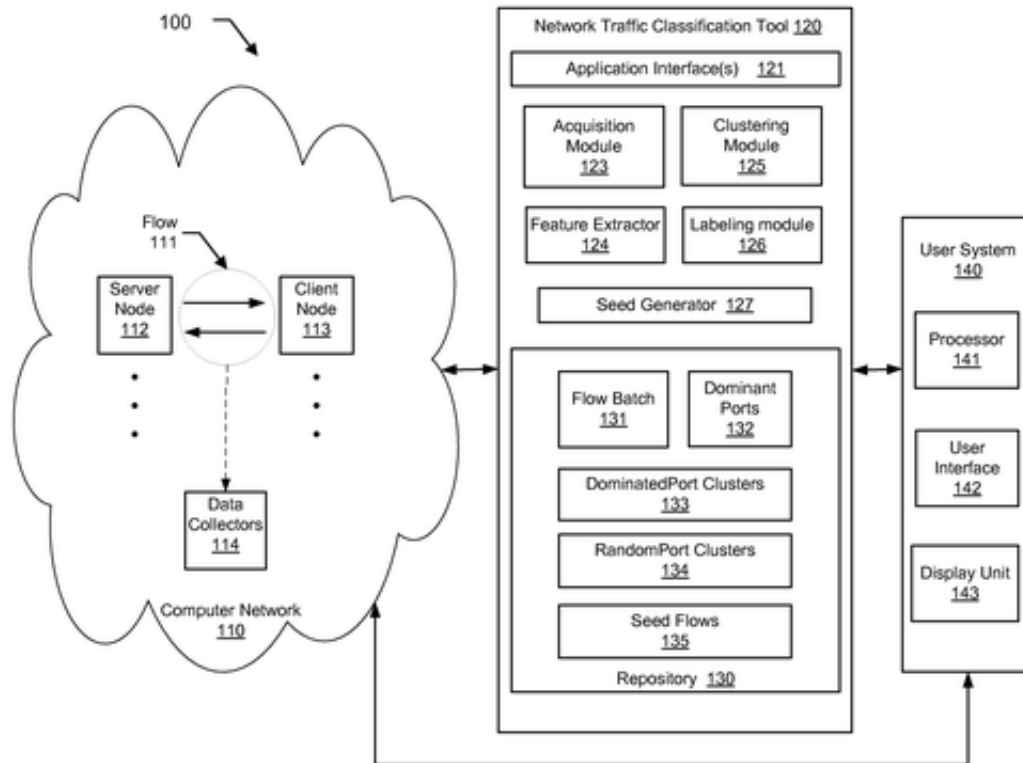
Filed by Ram Keralapura, San Jose, CA (US); Marco Mellia, Turin (IT); and Luigi Grimaudo, Turin (IT)

Filed on Nov. 18, 2011, as Appl. No. 13/300,342.

Int. Cl. G06F 15/173 (2006.01)

U.S. Cl. 709—224 [709/223]

24 Claims



1. A method for classifying network traffic in a network, comprising:
  - obtaining a first flow batch comprising a first plurality of flows from the network traffic;
  - processing, by a processor of a computer system, a first working set portion of the first flow batch for a first iteration based on a first pre-determined algorithm, comprising:
    - dividing the first working set portion into a plurality of clusters; and
    - filtering, based on a server port found in the cluster, a cluster of the plurality of clusters to generate a filtered cluster and a second working set portion of the first flow batch;
  - processing the second working set portion for a second iteration based on the first pre-determined algorithm; and
  - classifying the first flow batch based at least on the filtered cluster,
    - wherein filtering the cluster based on the server port comprises:
      - identifying a first server port as most frequently occurring comparing to all other server ports in the cluster;
      - in response to determining that a first frequency of occurrence of the first server port in the cluster exceeds a pre-determined threshold:
        - removing, from the cluster, a flow having a different server port than the first server port to generate the filtered cluster, wherein the filtered cluster is identified as a dominatedPort cluster based on a pre-determined criterion;
        - and
        - removing the dominatedPort cluster from the first working set portion to generate a remainder as a second working set portion.