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.