Service Usage Classification with Encrypted Internet Traffic in Mobile Messaging Apps

The rapid adoption of mobile messaging Apps has enabled us to collect massive amount of encrypted Internet traffic of mobile messaging. The classification of this traffic into different types of in-App service usages can help for intelligent network management, such as managing network bandwidth budget and providing quality of services. Traditional approaches for classification of Internet traffic rely on packet inspection, such as parsing HTTP headers. However, messaging Apps are increasingly using secure protocols, such as HTTPS and SSL, to transmit data. This imposes significant challenges on the performances of service usage classification by packet inspection. To this end, in this paper, we investigate how to exploit encrypted Internet traffic for classifying in-App usages. Specifically, we develop a system, named CUMMA, for classifying service usages of mobile messaging Apps by jointly modeling user behavioral patterns, network traffic characteristics and temporal dependencies. Along this line, we first segment Internet traffic from traffic-flows into sessions with a number of dialogs in a hierarchical way. Also, we extract the discriminative features of traffic data from two perspectives: (i) packet length and (ii) time delay. Next, we learn a service usage predictor to classify these segmented dialogs into single-type usages or outliers. In addition, we design a clustering Hidden Markov Model (HMM) based method to detect mixed dialogs from outliers and decompose mixed dialogs into sub-dialogs of single-type usage. Indeed, CUMMA enables mobile analysts to identify service usages and analyze end-user in-App behaviors even for encrypted Internet traffic. Finally, the extensive experiments on real-world messaging data demonstrate the effectiveness and efficiency of the proposed method for service usage classification.