STUDY ON ADVANCED TOPICS

Flow Extraction from mobile Data Packets

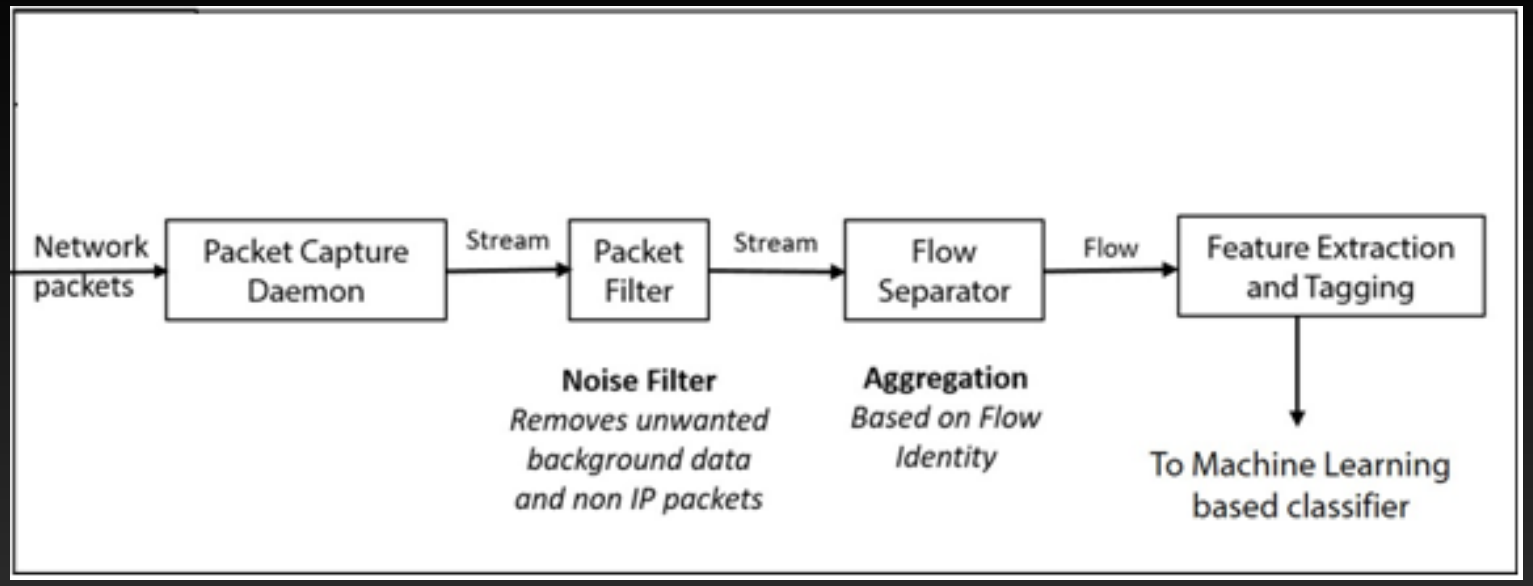
for the purpose of user activity / behaviour classification

Devamalya Hazra
Mohit Agarwal
Rishabh Sharma
Shubham Bansal

M.E. Software Systems
BITS Pilani, Department of Computer Science and Information Systems

Phases

- **Packet Capture**
- Packet Filtration
- Flow Separator
- Feature Extraction



Phases

- **Packet Capture**

  tcpDump with Android Debug Bridge (ADB)
  - Involves usage of ADB script.
  - Easier to tag in-app activities.
  - Not possible to simulate real usage.
  - Requires Root access of the device.

  TinyProxy
  - Intercepts network data at a centralized server
  - Accessible via sniffing utilities.
  - Feasible to scale.

  Requires isolation of the running app from the background activities to guarantee app-specific data packets.

- Packet Filtration
- Flow Separator
- Feature Extraction

Phases

- Packet Capture
- **Packet Filtration**
  - Removes unnecessary data.
  - Background data of proxy server
  - Non-IP packets like SSDP and ARP.
  - Used python scripts (scapy) to parse the .pcap files
- Flow Separator
- Feature Extraction

Phases

- Packet Capture
- Packet Filtration
- **Flow Separator**
  - Identify flow definition.
  - Identifies unique flows from the packet stream.
  - Flows to be associated with user activities in the later stages.
- Feature Extraction

Phases

- Packet Capture
- Packet Filtration
- Flow Separator
- **Feature Extraction**
  - Features identified and extracted from each aggregated flow
  - Size-based features
    - Concerned directly, or indirectly with the size of the flow
    - Number of packets
    - Average packet size
  - Direction-based features
    - A set of 20 features that deal with the direction of communication between the client and the server