Characterization of Individual Usage Patterns in Organizational Wireless Networks - A Case Study

Huijun Xiong Danfeng Yao
Department of Computer Science
Rutgers University, New Brunswick
Email: {huijun, danfeng}@cs.rutgers.edu

***Abstract*—Existing network analysis and security tools lack in two main aspects: *individualized analysis* and *personalized security*. Most current network trace analysis focuses on the aggregated traffic flow of the entire network, e.g., network-side traffic volume, busiest hosts on the network, and bursty periods of the organization. These types of network-wide traffic analysis do not give insights to the usage patterns of individuals on the network. It is desirable to identify hosts that have been behaving abnormally in comparison to their own historic records, which we call *personalized security*. We analyze the *individual usage patterns* in a wireless network of a university. We study both the statistical and temporal patterns of individuals' web usage behaviors from collected wireless network traces. Users are classified into different profiles based on their web usage patterns. We describe how our studies can be applied to improving network security, in particular detecting and preventing bot infection.**

Analyzing and characterizing organizational wireless network traces have traditionally been studied for maintaining the stability and availability of network resources. By studying the aggregated organizational network usages, system administrators aim to locate patterns of bursty traffic, local or area bottlenecks. Security is another important reason for these types of network analysis. It is reported that on average 3-5 percent of organizational assets are compromised by bots and malware - even when the best and most up-to-date security software is applied.

However, the existing network analysis and security tools lack in two main aspects: *individualized analysis* and *personalized security*. Most current network trace analysis focuses on the aggregated traffic flow of the entire network, e.g., network-side traffic volume, busiest hosts on the network, and bursty periods of the organization. These types of network-wide traffic analysis do not give insights to the usage patterns of individuals on the network. It is desirable to identify hosts that have been behaving abnormally in comparison to *their own* historic records, which we call the *personalized security* approach. However, network-wide traffic analysis is coarse-grained and insufficient in detecting *subtle* network attacks.

We analyze the individual usage patterns in a wireless network of a university. We analyze the *individual usage patterns* in a wireless network of a university. We study both the statistical and temporal patterns of individuals' web usage behaviors from collected wireless network traces. Users are classified into different profiles based on their web usage patterns.

Our characterization work ultimately aims to discover underlying patterns and properties of individual users' network behaviors. Our studies can be applied to improving network security, in particular to detecting bot infection. Botnet communication including command and control (C & C) and attacks disturbs the usual and routine traffic patterns of a user. Monitoring and comparing host-based traffic patterns with respect to the user's previous communication profile allow one to detect deviations and abnormalities caused by bots.

We obtained 4-month long wireless network traces from a university. The wireless network routinely serves three medium-sized department, one research center, and their visitors. We choose the top 500 users (represented by distinct local MAC addresses) who have highest number active days. We analyze all the traffic associated with HTTP protocol at port 80 of either the sender or the receiver and filter out the rest. Current bot C& C channels typically use HTTP at port 80.

We find the majority of users who visit a limited number of distinct websites. 432 users (out of the top 500 most active users) visited less than 100 distinct remote IP addresses during the 4 month-period, among them 278 users visited less than 50 IP addresses. Note that the number of distinct domains visited by a host may be even smaller than the number IP addresses, because IP addresses may belong to the same subnets and have the same domain suffixes.

We analyze how many *new* URLs and *old* URLs that a host visits each day. If an URL that a host visits exists in the surfing history, then it is labeled as an old URL, otherwise, new. We observe that most hosts visit fewer numbers of new URLs than old URLs each day, indicating that most nodes are somewhat consistent with their surfing history. There are several types of surfing dynamics in our dataset, for example, the *exploratory type* where a host constantly visits a significant number of new IP addresses each day, and the *bursty type* where a host demonstrates bursty new traffic occasionally, i.e., it occasionally visits many new websites. Hosts with the *conservative type* tend to visit old URLs in its surfing history.

For future work, we plan to construct more advanced pattern recognition techniques on the individuals' usage traces using machine learning methods. We also hope to analyze the how to quantify the similarity or difference in people's network usage behaviors.