WiFi Traffic Analysis Project Report.

Rajesh Golani, Srikanth Kandalam
Computer Science
Stony Brook University, NY

***Abstract*— WiFi Traffic analysis is a project which involved analysing the WiFi traffic in and around Stony Brook University. We present the details of the analysis and the results of our experiment on the WiFi using Network protocol analyser tool called Wireshark and a network traffic dump tool called Dumpcap. We have collected information about the WiFi in the form of packets and we have analysed the captured packets to understand the nature and performance of the WiFi at Stony Brook University.**

INTRODUCTION

Wi-Fi(Wireless Network LAN) has become one of the most prominent ways to connect all kinds of devices such as personal computers, audio players, tablets, smartphones and many kinds of digital devices. Any wireless local area network that follows IEEE 802.11 standard are considered as Wi-Fi. Wi-Fi has become a common terminology used by everyone but not many know about the intricate performance factors of a Wi-Fi network and how all the devices are able to stay connected using relatively few access points.

In this report, we present the results of our analysis which was done on the Wi-Fi traffic at Stony Brook University which helped us shed light on various interesting insights about the connectivity, performance and general traffic patterns of a wireless network. We have used Wireshark and Dumpcap utilities to analyse the Wi-Fi traffic. WiFi at Stony Brook University has 150 Mbps with IEEEE 802.11n protocol and is split into four networks namely 'WolfieNet-Secure', 'WolfieNet-Open', 'WolfieNet-Guest', 'WolifeNet-Get-Connected'. We have analysed the WiFi traffic networks at four different locations viz., Melville Library, Health Science Library, Computer Science Building, Chapin Apartments at the university.

This paper is organized in to multiple sections. 'Related Work' describes the work that has been previously done that is on similar lines of Wi-Fi traffic analysis. 'Packet Collection' section talks about the details of our sniffing such as the locations, time, amount of packet dump collected. "

RELATED WORK

David Kotz et al[1] have analysed the Wi-FI traffic at Dartmouth university for eleven weeks. The study was done on 476 access points and 161 buildings for 77 days. The paper presents a detailed analysis about the the amount of traffic a network can handle, amount of traffic per card and how the Wi-Fi traffic varies across the hours, days and weeks. It also concentrates on mobility of the users which led them find important aspects such as number of cards in the network, number of days a card is active and number of APs each card visit.

Mikhail Afanasyev et al[2] performed analysis on mixed usage of urban Wi-Fi network. They studied Google Wi Fi network at Mountain View, California for 28 days and found that the Wi Fi usage is identified under devices of three categories viz., Smart phone, Modem, Hotspot. Their research yielded the result that modem users are static and are always connected while placing highest demand on the network. Hotspot users are less number but concentrated only in few places like shopping malls and restaurants. Mobile category was, as expected, the one with the highest number of users and peak levels of activity.

PACKET COLLECTION

Background about the environment

Aruba Networks has installed 802.11ac Wi-Fi infrastructure in Stony Brook University 150 Mbps. 802.11ac is an extension introduced to 802.11n for continuing the thrust and extending the rates and throughput. It provides 1Gbps of maximum multi-station throughput and 500 Mbps of maximum single link throughput. Aruba Networks has installed 3000 Aruba 802.11n access points. On an average, every building will consist of 40-50 access points. Considering the increase in usage of smart phones, laptops by 24,000+ students and 17,000 staff university has upgraded the connection from 802.11n to 802.11ac. Wi-Fi his has been split widely across around 37 buildings in the campus with every building having it's own sub-divisions of the network. The four main sub-divisions of the Wi-Fi at our University are 'WolfieNet-Secure', 'WolfieNet-Open', 'WolfieNet-Guest', 'WolifeNet-Get-Connected'.
'WolfieNet-Secure' is a secure network which allows only college students, faculty and staff to access.
'WolfieNet-Open', 'WolfieNet-Guest',
'WolfieNet-Get-Connected' are open networks which allow everyone to access internet across the campus.

Packet collection details

Wireshark and dumpcap have been used to collect Wi-Fi packets across the following locations in the campus:

1) Melville Library*:* University library served as one of the hot spots for packet collection as approximately 15,000 students use the library weekly and hence there was a lot of Wi-Fi traffic. All kinds of devices like laptops, smart phones

etc... are used extensively inside the library and hence we have collected the network trace for 2 hour intervals on weekdays and weekends. This helped us analyse the traffic on the busiest days and normal days in the library.

2) Health Science Library*:* Health Science Library is located in Stony Brook Hospital which is another hotspot considering the fact that around 8000 - 9000 student use it weekly. Similar to Melville we have collected Wi-Fi packets on weekdays and weekends to understand the extremes.

3) Computer Science Building*:* Computer science building is located at the end of the campus and is relatively less populated compared to Melville and Health science libraries. Even though the density of access points in all the buildings is comparable, the usage of Wi-Fi by students is lesser in Computer Science building. Hence we would prefer to call it as one of our average traffic spot for our study.

4) Chapin Apartments*:* Chapin apartments are the residential buildings located near Stony Brook University and is occupied by graduate students. Among all the locations, Chapin is relatively a weak spot considering the fact that there are 11 blocks with approximately 10 apartments in each occupied by a maximum of 6 people which makes the total occupancy of 600-700. Also, there are approximately 100 access points for the whole apartments together. Hence we choose it to study the weaker extreme of the Wi-Fi traffic.

Methodology

Wireshark allows to collect packets which the network interface card is intended to receive. For better analysis, we required more data. Hence we operated Wireshark on *promiscuous* mode where it receives all the traffic without any discretion. This mode is vital for experiments which are intended to estimate network performance, throughput etc., because collecting packets which are intended only for our computer doesn't help us in anyway to determine those parameters accurately. We have achieved the promiscuous mode by creating an interface which interferes with the Network Interface Card and enables it to collect all the data in the air. The command to set this in Ubuntu is:

```
$sudo iw dev wlan0 interface add interfacename type monitor
```

And to bring up the interface, we have used to following command:

```
$sudo ifconfig interfacename up
```

Once we are done with collecting the packets, we have bring down the interface by using this command:

```
$sudo ifconfig interfacename down
```

Collection Time

We have sniffed the networks at all the four places viz., Melville Library, Health Science Centre, Computer Science building, Chapin Apartments on an hourly basis on 7 days on 4 different locations.

5) Melville and Health Science Centre: Sniffing at these locations was done on weekdays and they are the hotspots as there are a lot of students who access Wi-Fi at these places compared to other two places, CS Buildings and Chapin Apartments.

6) CS Building: Packets at CS building were collected during an off peak period (4 PM - 5PM) as we wanted to understand the behaviour of a hot spot at an off peak period which makes it a relatively weak spot.

7) Chapin Apartments: Sniffing at chapin was done at the peak period (10 PM - 11PM). This location is considered to be a weak spot as the maximum number of users at it's peak period is less than the average number of users at a hot spot.

RESULTS & ANALYSIS

In this section we present the results of our packet collection and the analysis of our results. The following are the basic statistics:

- Total number of days: 7
- Total number of access points covered: 347
- Total number packets collected: 6 million approx

Results at all locations

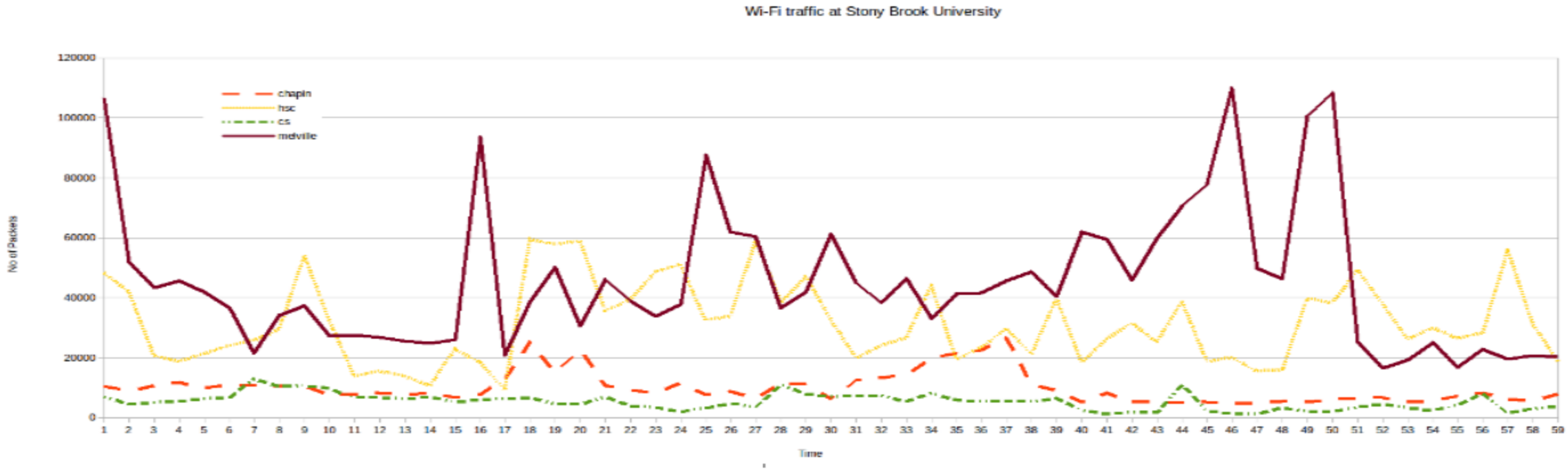
The graph below shows the trace of all locations sampled to one hour with time on X-axis and number of packets collected on Y-axis.

Analysis:

From the graph it is observed that:

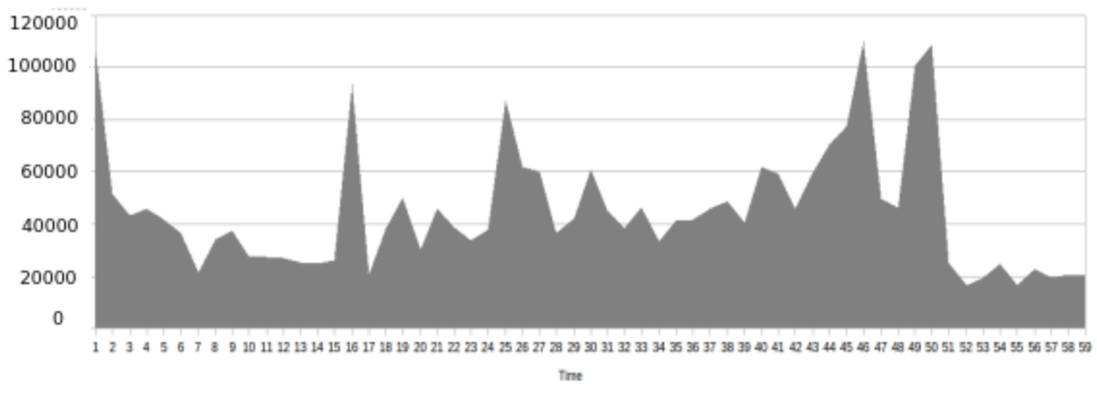
- Since Melville Library has the highest number of users (15,000) the graph demonstrates that highest traffic is collected here.
- Health Science Centre is another hotspot only next to Melville Library. The graph demonstrates the same. This location has the next highest traffic after Melville.
- Chapin Apartments peak traffic is still lower than both Melville and Health Science Centre and slightly higher than Computer Science building.
- Traffic at Computer Science Building represents the off-peak period and hence the lowest. Throughput is found to be .212Mbps.

Even though collection at Chapin apartments was at the peak time, the graph demonstrates the fact that the number of packets collected at Chapin apartments and the number of packets collected at CS building are comparable as the maximum number of users at Chapin (a weak spot) during it's peak time is still comparable to the minimum number of users at CS building during it's off-peak(a hot spot)



F. Results at Melville Library

- Throughput: 1.487 Mbps.
- Total number of access points covered: 133
- Total number packets collected: 2.5 million approx

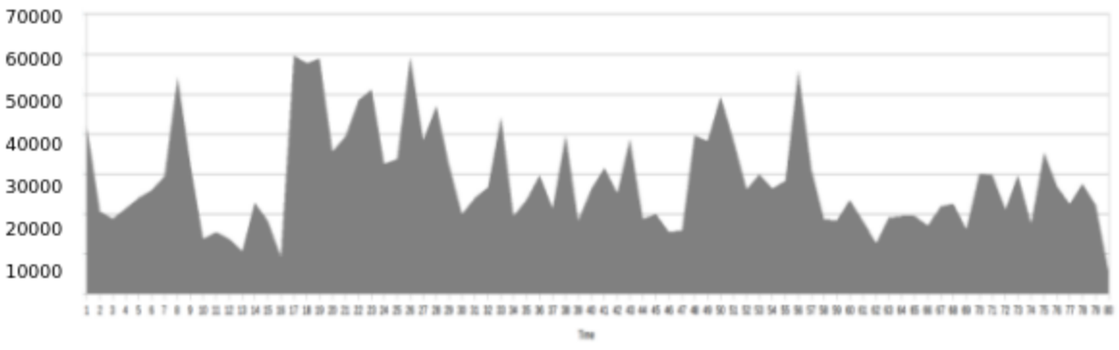


WiFi Traffic at Melville Library

X Axis: Time (in mins) Y Axis: Number of Packets

G. Results at Health Science Centre

- Throughput: 0.512 Mbps.
- Total number of access points covered: 77
- Total number packets collected: 1.5 million approx

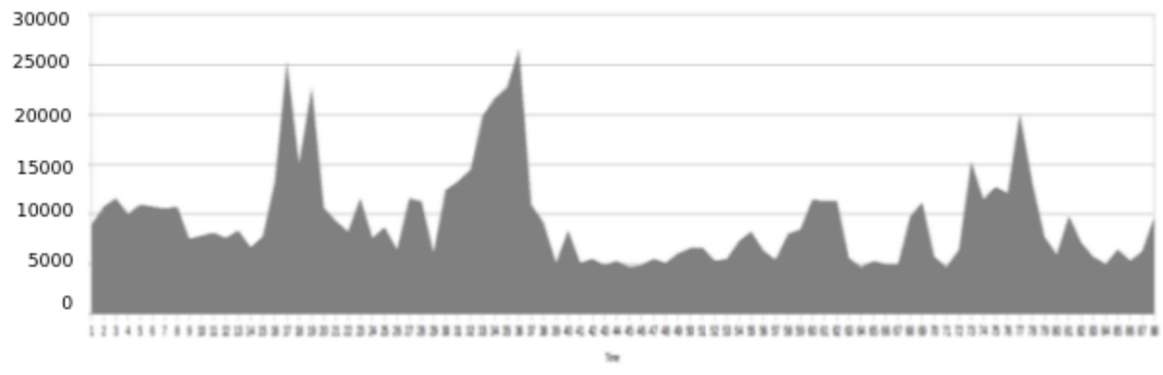


WiFi Traffic at Health Science Centre

X Axis: Time (in mins) Y Axis: Number of Packets

H. Results at Chapin Apartments

- Throughput: 0.512 Mbps.
- Total number of access points covered: 75
- Total number packets collected: 1 million approx

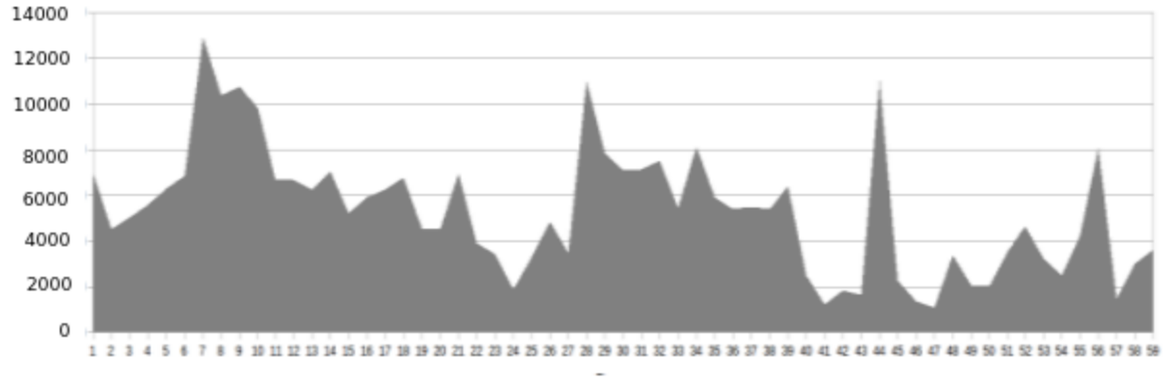


WiFi Traffic at Chapin Apartments

X Axis: Time (in mins) Y Axis: Number of Packets

I. Results at Computer Science Building

- Throughput: 0.212 Mbps.
- Total number of access points covered: 52
- Total number packets collected: 1 million approx



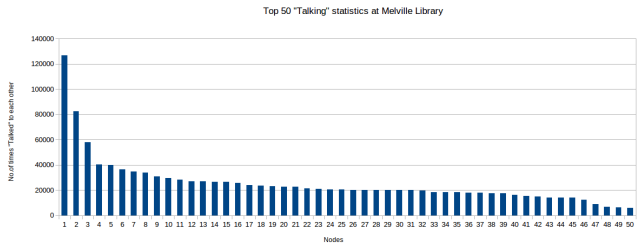
WiFi Traffic at Computer Science Building

X Axis: Time (in mins) Y Axis: Number of Packets

"Chatterbox pairs"

In every network the traffic is different for different nodes. Few nodes utilize it more while few utilize very less. For example, A had been spending a lot of time on internet downloading, playing games online etc., while B came checked his email and left. This results in A communicating a lot with access point than B. Let's call A and the access point together "Chatterbox pairs". Thus the more the number of "Chatterbox pairs", more Wi Fi traffic gets generated and thus the more the throughput of the network.
In technical terms, the pair who exchanges most number of packets can be termed as "Chatterbox pairs" We have analysed on similar lines and represented top 50 "Chatterbox pairs" of every location below:

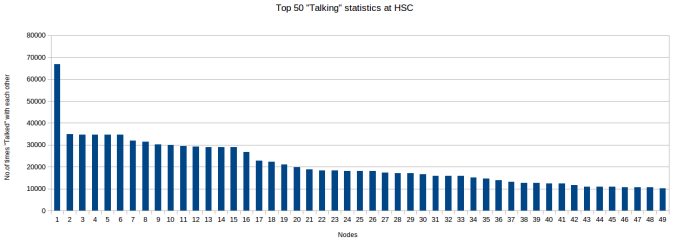
Melville Library: The throughput at Melville was found to be highest among all the four locations that we have analysed. Hence we can assume most number of "Chatterbox pairs" to be present here. The graph plotted for top 50 "Chatterbox pairs" proved the assumption:



Results

- The top "Chatterbox pair" of Melville recorded an exchange of 126594 packets in the span of 60 minutes.
- The average of top 50 "Chatterbox pairs" for Melville was found to be 25415.3 packets/hr

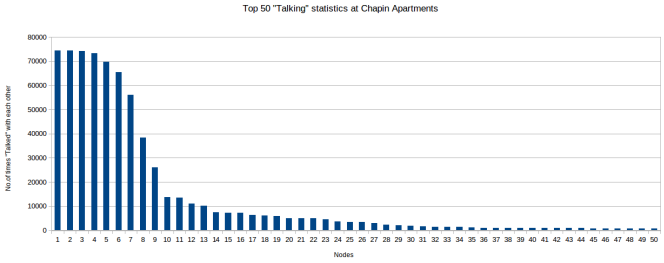
Health Science Centre: The throughput at Health Science Centre was found to be next to Melville library. Hence we expected quite a good number of "Chatterbox pairs" to be present here. The graph plotted for top 50 "Chatterbox pairs" proved the assumption:



Results

- The top "Chatterbox pair" of Health Science Centre recorded an exchange of 66832 packets in the span of 60 minutes.
- The average of top 50 "Chatterbox pairs" for HSC was found to be 20435 packets/hr

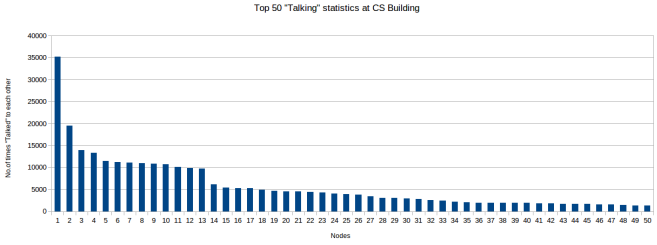
Chapin Apartments: Chapin Apartments was a weak spot and hence we weren't expecting "Chatterbox pairs" to be more aggressive than other locations. The graph plotted for top 50 "Chatterbox pairs" proved the assumption:



Results

- The top "Chatterbox pair" of Chapin recorded an exchange of 74402 packets in the span of 60 minutes. Even though the top "Chatterbox pair" at Chapin is slightly higher than HSC, there were very few pairs close to that value. And from the graph it can be observed that the after top 6 users, the value fell sharply which proves that there were very few active users at Chapin apartments.
- The average of top 50 "Chatterbox pairs" for Chapin was found to be 14834 packets/hr

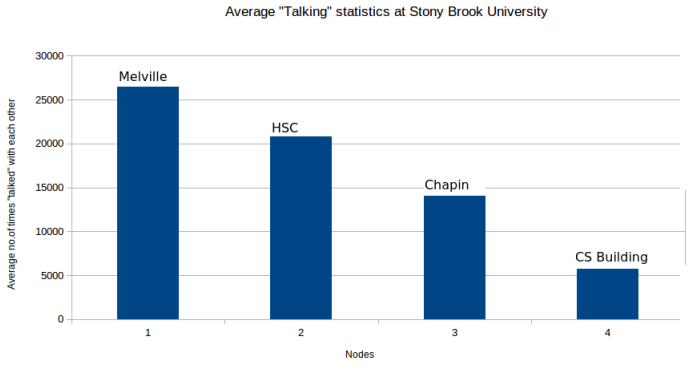
Computer Science Building: Computer Science was the weakest spot of all and hence we assumed very few "Chatterbox pairs" to be present with a low average. The graph plotted for top 50 "Chatterbox pairs" proved the assumption:



Results

- The top "Chatterbox pair" of CS building recorded an exchange of 35197 packets in the span of 60 minutes.
- The average of top 50 "Chatterbox pairs" for CS building was found to be 5902 packets/hr

The average ”Chatterbox pairs” of all the locations is represented in the graph below:

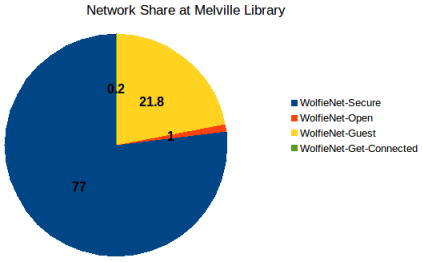


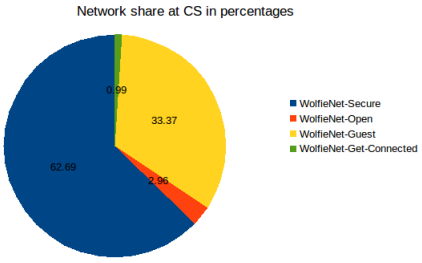
Most sought after Network

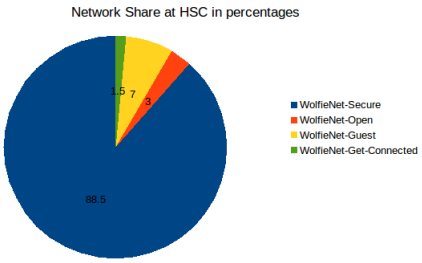
At Stony Brook University, Wi Fi is divided into 4 main sub divisions: *WolfieNet-Secure, WolfieNet-Open, WolfieNet-Guest and WolfieNet-Get-Connected*. Every department has it’s own sub divisions apart from these 4 main sub divisions. But we have concentrated only on these 4 sub-divisions for simplicity.

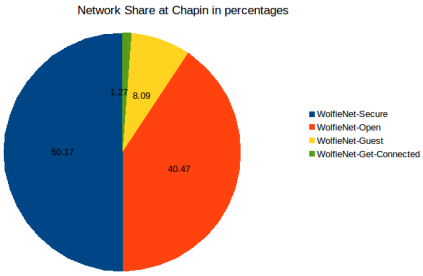
As stated at the beginning of this paper, to get connected to *WolfieNet-Secure*, you need to be a faculty/student/staff of the university and for *WolfieNet-Open, WolfieNet-Guest and WolfieNet-Get-Connected* you don’t need a password. Hence we wanted to analyse how many people are preferring to connect to the secure network (*WolfieNet-Secure*) and how many are not at different locations. This is an indicator of how easy it is to attack the network and might help us get a perspective of educating people to use secure networks for all sensitive transactions.

For this analyses, we have filtered ’Probe Requests Frames’ which are essentially the requests sent by the user to obtain information from either an access point specified by SSID or all access points in the area, specified with the broadcast SSID. Each Probe Request frame SSID associated with it. Hence if user is connected to *WolfieNet-Secure* and trying to communicate with it, it will send a Probe Request with SSID = ”WolfieNet-Secure”. Hence this way we were able to find the count of each network at every location and represented the results in the graphs below:









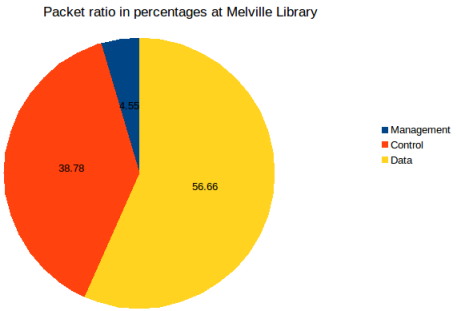
From the above graphs, we could understand that most of the people are using *WolfieNet-Secure* and the next most dominating network is *WolfieNet-Guest*. *WolfieNet-Guest* is an open network which can be accessed without any password. This was mainly intended for the people who visit the campus. Hence it could have both good and bad consequences.

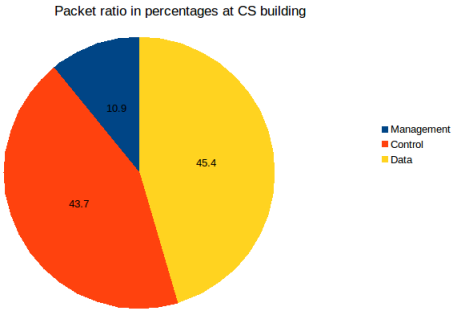
Content of the packets

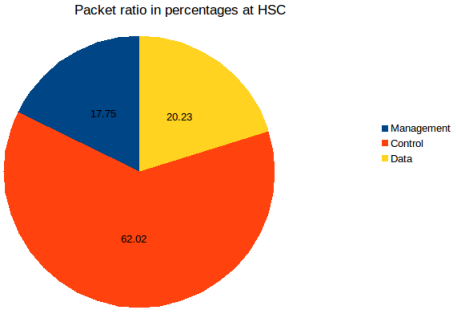
After collecting approximately 6 million packets from all the locations, we have analysed the content of the packets. We wanted to know how much percentage of packets are data packets, management packets and control packets because it would help us understand how much of the network is really being used for data transfer and how much is being flooded with Broadcast, Beacon, Probe frames etc. From a capture file, we can filter Management, Control, Data packets using these filters in Wireshark:

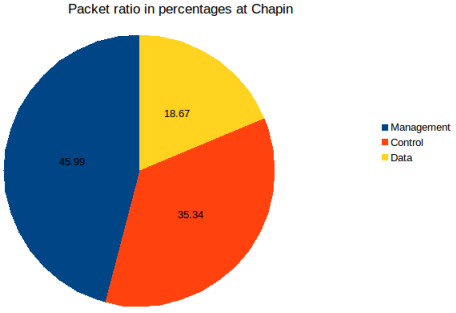
- Management - wlan.fc.type eq 0
- Control - wlan.fc.type eq 1
- Data - wlan.fc.type eq 2

We have plotted the results below for each location:









From the graphs, only at Melville and CS building there is a fair share for Data packets while in the other locations, it has been restricted to 18-21 percent. This result could be interpreted in multiple ways. CS building having the lowest number of users makes it possible to have more data packets share as there wouldn't be much scope to send Beacon frames/Probe frames/Association requests which are part of Management and Control packets. But at the same time Melville Library has the highest number of users yet it has high data share. How can we explain this? We think this was possible because of more number of access points are located at Melville library(77) compared to CS building (52) which makes it easier to manage more number of users.

REFERENCES

[1] Characterizing Usage of a Campus-wide Wireless Network David Kotz and Kobby Essien Department of Computer Science, Dartmouth College .

[2] Analysis of a Mixed-Use Urban WiFi Network: When Metropolitan becomes Neapolitan Mikhail Afanasyev, Tsuwei Chen , Geoffrey M. Voelker, and Alex C. Snoeren University of California, San Diego and Google Inc.