Large-scale Measurements of Wireless Network Behavior

Sanjit Biswas
biswas@samsara.com

John Bicket
jbicket@samsara.com

Edmund Wong
elwong@meraki.com

Raluca Musaloiu-E
ralucam@meraki.com

Apurv Bhartia
apurv@meraki.com

Dan Aguayo
aguayo@meraki.com

Cisco Meraki
500 Terry Francois Blvd.
San Francisco, CA 94158

ABSTRACT

Meraki is a cloud-based network management system which provides centralized configuration, monitoring, and network troubleshooting tools across hundreds of thousands of sites worldwide. As part of its architecture, the Meraki system has built a database of time-series measurements of wireless link, client, and application behavior for monitoring and debugging purposes. This paper studies an anonymized subset of measurements, containing data from approximately ten thousand radio access points, tens of thousands of links, and 5.6 million clients from one-week periods in January 2014 and January 2015 to provide a deeper understanding of real-world network behavior.

This paper observes the following phenomena: wireless network usage continues to grow quickly, driven most by growth in the number of devices connecting to each network. Intermediate link delivery rates are common indoors across a wide range of deployment environments. Typical access points share spectrum with dozens of nearby networks, but the presence of a network on a channel does not predict channel utilization. Most access points see 2.4 GHz channel utilization of 20% or more, with the top decile seeing greater than 50%, and the majority of the channel use contains decodable 802.11 headers.

CCS Concepts

•Networks → Wireless access points, base stations and infrastructure; Network measurement;

SIGCOMM '15, August 17 - 21, 2015, London, United Kingdom

DOI: http://dx.doi.org/10.1145/2785956.2787489

Keywords

802.11, large-scale measurements, network usage data

1. INTRODUCTION

Over the past 20 years, wireless LANs based on 802.11 have become common in office and campus environments. Recent estimates suggest over 10 billion WiFi devices have been sold in total and that over 4.5 billion of those devices are in use today [3]. These devices use the same underlying standards and frequency bands defined in the mid-1990s, resulting in an increasingly crowded wireless environment.

Several papers have studied wireless network behavior, from indoor and outdoor links, to campus-scale user behavior. These papers have helped improve our understanding of how real-world networks behave, resulting in improvements to simulators, debugging tools, and protocols. However, there are few studies that analyze a large cohort of wireless LAN networks or over a long period of time.

The Meraki system provides a unique perspective over a wide range of networks due to its cloud-based management architecture. Each Meraki access point is polled periodically by the Meraki system for a number of statistics about the state of the wireless channels, associated clients, and application traffic statistics for each associated client. This information is written to a database, aggregated, and presented to network administrators through a central web service.

Since its inception in 2006, this database has grown to contain information from millions of network devices, including access points, switches, routers, and client devices such as laptops and smartphones.

This paper contributes a detailed look at network behavior, using an anonymized subset of the data contained in the Meraki system. By filtering for devices based on their type of hardware platform and choosing a large set of devices, we are able to plot trends which are not influenced by factors such as variations between chipsets, changes in customer type, or network policy. Specifically, we look at the the following four sets of data:

	Cisco Meraki MR16	Cisco Meraki MR18
CPU	Qualcomm Atheros AR7161 680MHz	Qualcomm Atheros QCA9557 SoC
Memory	64MB DDR	128MB DDR2
Radio	Qualcomm Atheros AR9223 (2.4 GHz), AR9220 (5 GHz) PCI 2x2 802.11n chipset	On-chip Qualcomm Atheros QCA9557 radio (2.4 GHz) with external power amplifier, AR9592 5 GHz PCI 2x2, AR9592 1x1 PCI scanning radio
Transmit power	23 dBm (2.4 GHz), 24 dBm (5 GHz)	24 dBm (2.4 GHz), 24 dBm (5 GHz)
Antenna	Built-in 3dBi 2.4 GHz, 5dBi 5 GHz	Built-in 3dBi 2.4 GHz, 5dBi 5 GHz

Table 1: Hardware platforms used to measure link delivery and channel utilization in Sections 4 and 5

1. For a set of 20,667 networks which serve over five million unique clients a week, we examine the wireless capabilities of the clients as well as the OSes and applications the clients use. Our data also provides insight into the traffic characteristics of various applications as well as how clients have changed in the past year.

2. For a set of 10,000 Cisco Meraki MR16 802.11n access points, we summarize the link delivery characteristics of over 20,000 wireless link pairs measured by periodic broadcasts sent by the access points.

3. Using the same set of access points, we examine the average channel utilization of the 2.4 GHz and 5 GHz bands over a period of six months.

4. With a set of 10,000 Cisco Meraki MR18 802.11n access points, which contain a third radio capable of capturing spectral data, we examine more detailed short-term channel measurements.

The remainder of the paper is organized as follows: Section 2 outlines the architecture of the Meraki system and details about the hardware platforms used in the study. Section 3 studies client and application usage and provides tables demonstrating changes in usage over several years. Section 4 describes our method for measuring link-level statistics between access points and studies the distribution of link delivery rates. Section 5 looks specifically at wireless channel utilization at a fine-grain level. Section 6 describes some real-world challenges faced by the Meraki system. Finally, Section 7 discusses related work, and Section 8 concludes.

2. SYSTEM ARCHITECTURE

The Meraki system consists of access points, switches, and firewalls at customer sites, and a centrally hosted management system, known as the backend system.

Each network in the Meraki system can contain a combination of these wireless and wired network devices that all report data to the backend system in the same way. The backend system itself is distributed across several data centers; for the purposes of this paper, we treat the backend system as a single data store.

Each piece of Meraki networking equipment maintains persistent encrypted tunnels to two different backend data centers. Using these connections, the backend periodically harvests statistics from each network device using protocols, built with Google Protocol Buffers [21] to minimize reporting overhead.

A typical access point averages around 1 kilobit per second to report to the backend. These tunnels are used only for statistics and configuration; client traffic is routed directly to the local network or the Internet. In the event a device is unable to reach the Meraki backend, normal client routing and accounting continues. The backend polls for queued information when the connection is reestablished.

The Meraki backend system is designed to handle machine upgrades, schema changes and new software revisions without affecting the measurement data. The system operates using a pull mechanism, which helps regulate the flow of updates to the database during times of peak load.

2.1 Hardware platforms

The various Meraki hardware platforms contain a similar system architecture and share a Linux codebase, but each product family contains application-specific chipsets. They consist of a general purpose CPU (single or multi-core) running Linux and the Click Modular Router [14], coupled with multiple 802.11 radios (for wireless access points), multiple Gigabit Ethernet interfaces (for security appliances) or a 24/48 port Gigabit or 10 Gigabit switch fabric (for edge switches). This paper focuses only on data from wireless access points.

Each the platform is designed to forward traffic at line-rate while still being able to track client metadata and detailed application usage statistics about individual TCP flows. The fast data path is handled either in hardware (for switching platforms) or in the device driver and in the Click modular router (for security appliances and wireless access points). Elements within the Click modular router on the fast path handle policy routing decisions, client firewall rules, and track application classification and usage for each MAC address. Other specific types of traffic are processed along the slow path, such as ARP, DHCP, DNS, multicast DNS, TCP SYN/FIN, packets containing HTTP headers, and packets containing SSL handshakes. Traffic along the slow path is filtered in the driver or hardware and handled entirely by the Click modular router to identify specific application traffic. Click contains elements to extract additional in-memory

metadata information from these traffic flows, such as hostnames, operating system fingerprints, and application-level usage data. This metadata is used to identify flows and update application usage data counters, similar to [17]. There are about 200 application identification rules that use the metadata to identify TCP flows and applications. The user and application analysis was done on a worldwide set of 20,667 networks containing two Meraki access point models, which are further described in Section 3.

The wireless measurements in Section 4 were taken on set of 10,000 identical Cisco Meraki MR16 access points located in the US to simplify analysis. These access points were deployed in between 2010 and 2014 were in continuous operation measurement period from January 2014 through January 2015. Similarly, in Section 5 we consider a set of 10,000 identical Cisco Meraki MR18 access points, equipped with a dedicated scanning radio. All access points were located in the United States and followed the FCC Part 15 regulatory limits. Table 1 describes the access point hardware used to create the data set.

2.2 Firmware revisions

During the measurement period, there were a total of 2 major firmware revisions applied to the access points. The rough dates of these upgrades are January and December 2014. The lower layers of the driver and HAL were not changed, as the updates included routine package security updates and added higher level features related to management and access control features.

2.3 Data collection

In addition to per-client and flow traffic statistics, each access point measures and records wireless channel statistics. The traffic statistics and additional measurements are periodically polled by the Meraki backend system and recorded in long term storage. For usage statistics such as byte counters and application usage, local statistics are aggregated by MAC address in the backend (to account for roaming).

3. NETWORK USAGE

In this section, we take a look at the clients that connect to our access points. In particular, we examine client radio capabilities, the OSes and applications that are most commonly used by clients, and how usage patterns have changed over the past year.

The data presented here were collected from 20,667 wireless networks operated by 11,788 different administrative organizations. All of these networks have at least two wireless access points and have application traffic profiling enabled. All usage data was collected over the one-week period of January 15-22 in 2014 and 2015. We show the numbers for 2015 and provide the percent increase ("% increase") compared to 2014.

The sampled networks include both urban and rural deployments throughout the world. Table 2 describes the mix of networks in the data set organized by industry type.

To preserve anonymity, all of our data are presented only as an aggregate over all of these networks.

Industry	# networks
Architecture/Engineering	127
Construction	333
Consulting	365
Education	4,075
Finance/Insurance	737
Government/Public Sector	1,112
Healthcare	1,382
Hospitality	493
Industrial/Manufacturing	1,220
Legal	264
Media/Advertising	427
Non-Profit	640
Real Estate	386
Restaurants	296
Retail	2,355
Tech	983
Telecom	442
VAR/System Integrator	2,876
Other	2,154
Total	20,667

Table 2: Network deployment types for the application usage data set. There are a wide range of industrial verticals and the networks are not dominated by one particular industry.

3.1 Client devices and signal strength

Our access points collect information about the 802.11 capabilities advertised by every client that connects to the network. Table 4 summarizes how these capabilities have changed over one year.

In summary, we found that:

- 802.11ac clients became much more common over the year: at the end of the sampling period 18% of clients were 11ac-capable.
- A majority of clients now have 5 GHz capability, but nearly 40% of all clients remain 2.4 GHz only.
- Multi-stream clients, became more common, and now about 25% of clients support multiple spatial streams.

Another item of interest is the strength of clients' connections to the network. Our system does not keep a historical record of client signal strength, but, through our centralized infrastructure, we are able to collect a snapshot of RSSI across all connected clients in real time. Figure 1 shows the distribution of client signal strength among about 309,000 clients that were connected one evening in January 2015.

At the time of the snapshot, about 249,000 (80%) of these clients were connected on the 2.4 GHz band, and 60,000 were connected at 5 GHz. This is interesting given the above data which show that about 65% of clients are 5 GHz capable. The difference is presumably due to greater attenuation at 5 GHz, which makes it harder for clients to associate on the higher band.

OS	TB (% total/% download)	% increase	# clients	% increase	MB / client	% increase
Windows	589 (30%/83%)	43%	822,761	28%	751	12%
Apple iOS	545 (28%/88%)	92%	2,550,379	34%	224	44%
Mac OS X	445 (23%/75%)	44%	313,976	24%	1,487	17%
Android	177 (9.1%/89%)	172%	1,535,859	61%	121	69%
Unknown	78 (4.0%/45%)	-9.2%	228,182	-8.9%	357	-0.36%
Chrome OS	62 (3.2%/91%)	275%	178,095	222%	366	16%
Other	26 (1.3%/78%)	80%	13,969	-33%	1,951	168%
Sony Playstation OS	22 (1.1%/96%)	53%	4,267	-13%	5,319	77%
Linux	5.8 (0.30%/68%)	611%	4,402	165%	1,393	169%
RIM BlackBerry	0.14 (0.0074%/94%)	-62%	13,681	-53%	11	-19%
Mobile Windows OSes	0.12 (0.0064%/91%)	-35%	4,943	-42%	26	13%
All	1,950 (100%/82%)	62%	5,578,126	37%	367	18%

Table 3: Usage by operating system during January 15-22, 2015. "% download" is the percentage of download (vs. total) traffic for a particular OS. "% increase" reflects the year-over-year increase from January 15-22, 2014.

	Jan. 2014	Jan. 2015
802.11g	99.9%	99.9%
802.11n	95.7%	97.7%
5 GHz	48.9%	64.9%
40 MHz channels	23.4%	63.8%
802.11ac	2.5%	18.0%
Two streams	7.7%	19.3%
Three streams	2.4%	3.8%
Four streams	0.7%	1.8%

Table 4: Client capabilities advertised by all clients that connected during the same week in January for two consecutive years.

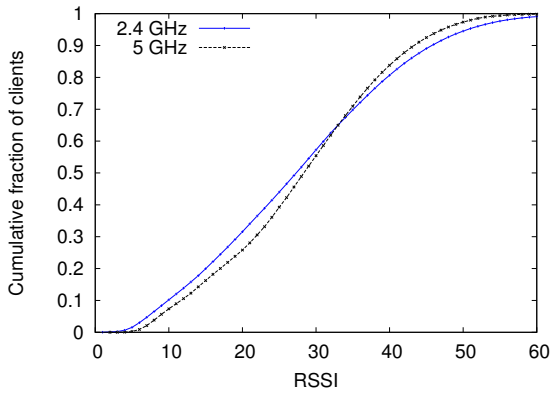

Figure 1: Distributions of received signal strength (RSSI) as measured at the access points, for approximately 309,000 associated clients.

This last point was corroborated by the signal strengths we observed. Median signal strength among all clients was about 28 dB above the noise floor on both bands but was typically lower on 5 GHz connections, again presumably because of attenuation.

3.2 Usage by operating system

Meraki uses a combination of MAC address prefix, DHCP fingerprints, and HTTP User-Agent inspection to determine device types, similar to methods described in [5]. Table 3 summarizes the breakdown of clients by operating system with their total bandwidth usage during a one-week period between January 15 and 22 in both 2014 and 2015. The unknown OS row represents clients where our heuristics could not classify the device, such as with devices running virtual machines or dual-boot operating systems (which would present multiple DHCP fingerprints from a single MAC address) or with embedded devices running Linux. More common is multiple device types presented by the user agent string, sent by browsers such as Chrome or smartphone applications. The reduction in unknown devices between January 2014 and 2015 is due to improvements in our heuristics.

Overall, we observed the total population of clients grow by approximately 37% from 4.07 million to 5.58 million devices year-over-year. Total usage grew by roughly 62%, an increase of 18% per client on average.

Because many OSes are specific to particular devices, Table 3 provides insight into usage by both OS and device type. As mentioned earlier, the classification of device types improved between the two measurement periods, which makes it difficult to compare device type growth in absolute terms. We can, however, compare average usage per client, which shows an steep increase of approximately 69% and 44% for Android and Apple iOS devices, versus a more moderate increase of 12% and 17% for Windows and Mac OS X devices. Even with these increases, mobile devices generally consume at most a few hundred MB of data on average in a week, whereas laptops and desktops running Windows and Mac OS X consume several times more (in particular, Mac OS X devices consumed roughly twice the bandwidth that their Windows counterparts did).

Our data also shows that clients using mobile devices tend to use a larger fraction of their bandwidth consuming con-

tent as compared to traditional desktops and laptops. Mobile devices download roughly 9 times more than they upload (in contrast, Mac OS X devices download closer to 3 times more). While mobile platforms have lower average usage, the number of mobile devices greatly outnumber laptops and desktops: we saw three times more Apple iOS devices than Windows devices, making the total usage of Apple iOS comparable to Windows, the most popular OS by usage.

Linux also registered extraordinary growth on our networks, which we attribute to the small number of clients in 2014, improved OS identification heuristics, and the increasing proliferation of Linux-based embedded devices. Finally, we observed significant per-client usage for Sony Playstation and Microsoft Xbox (not shown), most of which is downstream. This usage is not surprising given that game consoles are often used to stream media and play games that use network connectivity for content or functionality.

3.3 Application usage

As described in Section 2.3, Meraki uses several sources of information—including initial DNS lookup, HTTP header inspection, SSL handshake inspection, and port numbers—to determine the application underlying each new network flow. These periodically-updated fingerprints are applied as rule sets within the Click router running on each access point to simplify statistics collection.

Table 5 shows the top 40 applications by bytes transferred (both upstream and downstream) across the same 5.58 million clients as in Section 3.2. The number of clients is displayed along with the usage for a given application. Several of the categories—including miscellaneous web, miscellaneous secure web, miscellaneous video, miscellaneous audio, non-web TCP, and UDP—capture flows from applications not described in the rule set. We also classify each application into a category and show the total usage of various application categories in Table 6. While intuitive, our client usage data enables us to quantify how much traffic various applications and application categories consume in our networks, which applications have grown in the past year, and how individual clients participate in these services. The overhead of data collection was less than 1 kbit/second per access point on average and does not result in enough usage to register in the top 40 applications table.

Video and music applications such as YouTube, Netflix, iTunes, Spotify, and Hulu make up the largest fraction of usage at 34%, with 97% of their usage being download traffic. Looking specifically at Netflix, we found that each client consumed nearly 1.2 GB in a week. These services also recorded significant increases in both the total usage and number of clients that used these services.

File sharing—within a LAN (e.g., Windows file-sharing) and through cloud-based services (e.g., Dropbox)—was the second largest category at 8.4% of total usage with overall growth of approximately 28% year-over-year. Interestingly, we found that content uploaded using file-sharing services is not shared widely on average, as file-sharing services have a very balanced download/upload profile overall. The popularity of file sharing and its balanced downstream/upstream characteristics stand in contrast with web file sharing (services that distribute files via web links, e.g., mediafire.com and hotfile.com). While superficially similar, web file-sharing clients use nearly 45.4 more downstream bandwidth than upstream, implying that content is often uploaded once but downloaded multiple times. Online backup, another similar type of application, is on the other extreme: clients upload 22.8 times more data than they download, as restoring from backup is a relatively rare operation. Overall, our networks generally experience about 4.6 times more downstream traffic than upstream, with VoIP and video conferencing as well as online backup being the rare exceptions of application categories that use more upstream bandwidth than downstream.

One notable data point among the top 40 is Dropcam, a WiFi video-streaming camera and associated cloud backend service for storing and watching the resulting video. Dropcam has the fewest clients (2,940) among the top 40, less than a quarter of the number of clients that the next smallest application has. Yet, each client uses roughly 2.8 GB a week and uploads nearly 19 times more than they download, implying that Dropcam users do not often watch what they record. The extraordinary amount of usage puts Dropcam among the top 30 applications that we observed, above many applications that have far more clients and above several video, P2P, and file-sharing applications, categories that are traditionally heavy hitters.

4. INTERFERENCE LEVELS

As WiFi, Bluetooth, and other unlicensed spectrum devices have gained in popularity, there has been increased interference from nearby devices. In this section, we measure both the number of nearby WiFi access points as well as the amount of time the energy detect mechanism is triggered (a measure of channel utilization), in an effort to understand how busy unlicensed spectrum is in practice.

4.1 Nearby networks

Each Meraki access point is equipped with multiple radios: one dedicated to the 2.4 GHz ban and one dedicated to 5 GHz band (channels selectable based on regulatory region). The newer Meraki MR18 platform also has a third radio that scans both bands simultaneously. The radios are capable of decoding 802.11 a/b/g/n signals in both 20 MHz and 40 MHz formats, making it possible to decode frames from several generations of nearby access points. The radios are not capable of decoding other types of beacons from Bluetooth, Zigbee, or other standards; these beaconds are detected as noise during spectrum analysis.

As with the Meraki radios, many new access points are capable of simultaneous operation in the 2.4 GHz and 5 GHz bands. In most cases, access points send out a Broadcast SSID (BSSID) beacon frame every 102.4 ms (the default BSSID beacon interval) for 0.42 ms in the case of 802.11 a/g/n beacons and 2.592 ms for 802.11b beacons. As the number of nearby SSID broadcasts increase, the beacons and associated data frames may trigger the local AP's car-

Application	Category	TB (% total / % download)	% increase	# clients	% increase	MB / client	% increase
Miscellaneous web	Other	239 (13%/87%)	67%	4,623,630	36%	54	23%
YouTube	Video & music	202 (11%/97%)	93%	1,934,371	53%	110	26%
Netflix	Video & music	188 (9.8%/98%)	76%	161,014	19%	1,224	48%
Non-web TCP	Other	156 (8.2%/64%)	51%	3,656,494	40%	45	8.3%
Miscellaneous secure web	Other	147 (7.7%/70%)	94%	5,115,023	40%	30	38%
iTunes	Video & music	102 (5.4%/98%)	66%	2,230,787	38%	48	20%
Miscellaneous video	Video & music	98 (5.1%/91%)	61%	1,383,386	76%	74	-8.6%
Windows file sharing	File sharing	87 (4.5%/66%)	48%	740,591	31%	123	13%
CDNs	Other	75 (3.9%/72%)	81%	3,157,028	46%	25	24%
UDP	Other	61 (3.2%/61%)	60%	3,705,171	69%	17	-5.6%
Facebook	Social web & photo sharing	53 (2.8%/92%)	127%	3,579,926	35%	16	68%
Google HTTPS	Other	49 (2.6%/85%)	67%	3,953,002	44%	13	16%
Apple file sharing	File sharing	42 (2.2%/44%)	18%	21,951	-1.7%	2,005	21%
apple.com	Other	37 (1.9%/94%)	79%	2,763,663	32%	14	36%
Google	Other	34 (1.8%/85%)	19%	3,804,317	39%	9.5	-14%
Google Drive	Other	24 (1.2%/79%)	374%	1,325,938	138%	19	99%
Dropbox	File sharing	23 (1.2%/60%)	-1.5%	369,068	29%	65	-24%
Software updates	Software & anti-virus updates	18 (0.94%/98%)	36%	689,677	16%	27	18%
Instagram	Social web & photo sharing	17 (0.91%/96%)	45%	831,935	50%	22	-3.3%
BitTorrent	P2P	13 (0.69%/58%)	-8.5%	38,294	15%	360	-21%
Skype	VoIP & video conferencing	13 (0.69%/49%)	48%	392,878	27%	35	16%
Miscellaneous audio	Video & music	13 (0.66%/97%)	54%	460,262	60%	29	-3.7%
Pandora	Video & music	12 (0.64%/97%)	25%	182,753	34%	70	-6.8%
RTMP (Adobe Flash)	Other	12 (0.62%/96%)	10%	141,403	6.2%	88	3.8%
Gmail	Email	12 (0.62%/74%)	26%	1,337,755	42%	9.2	-11%
microsoft.com	Other	11 (0.59%/94%)	15%	861,136	34%	14	-15%
Tumblr	Other	11 (0.57%/97%)	31%	270,482	21%	42	7.9%
Spotify	Video & music	11 (0.56%/98%)	142%	209,219	115%	54	13%
Windows Live Hotmail and Outlook	Email	9.0 (0.47%/64%)	216%	366,272	108%	26	52%
Dropcam	VoIP & video conferencing	8.0 (0.42%/5.0%)	72%	2,940	155%	2,836	-32%
Hulu	Video & music	6.9 (0.36%/98%)	102%	51,667	100%	140	0.90%
Steam	Gaming	6.6 (0.35%/98%)	47%	21,011	45%	332	1.4%
Twitter	Social web & photo sharing	6.4 (0.33%/91%)	67%	1,925,505	34%	3.5	25%
Encrypted P2P	P2P	6.3 (0.33%/97%)	17%	81,673	23%	81	-4.5%
Encrypted TCP (SSL)	Other	6.0 (0.31%/65%)	50%	1,441,775	49%	4.3	0.81%
Remote desktop	Other	5.5 (0.29%/88%)	66%	93,876	13%	61	46%
ESPN	Sports	5.1 (0.27%/98%)	122%	202,971	41%	27	58%
Xfinity TV	Video & music	4.9 (0.26%/98%)	87%	12,802	27%	400	47%
Other web-based email	Email	4.7 (0.25%/49%)	-6.4%	277,919	23%	18	-24%
Microsoft Skydrive	File sharing	4.4 (0.23%/25%)	-10%	269,437	12%	17	-20%

Table 5: Top 40 applications by usage during January 15-22, 2015. See Table 3 for explanations on columns.

Category	TB (% total/% download)	% increase	# clients	% increase	MB / client	% increase
Other	901 (47%/77%)	65%	5,617,395	39%	168	19%
Video & music	648 (34%/97%)	75%	5,047,976	49%	135	17%
File sharing	160 (8.4%/58%)	28%	1,209,821	28%	138	0.12%
Social web & photo sharing	81 (4.2%/93%)	92%	4,691,155	39%	18	38%
Email	32 (1.7%/67%)	41%	2,632,542	38%	13	2.5%
VoIP & video conferencing	24 (1.3%/35%)	55%	483,222	22%	52	27%
Peer-to-peer (P2P)	20 (1.0%/70%)	-1.3%	113,720	22%	184	-19%
Software & anti-virus updates	20 (1.0%/98%)	28%	699,776	15%	29	11%
Gaming	11 (0.57%/96%)	49%	199,804	45%	57	2.8%
Sports	5.3 (0.28%/98%)	117%	225,875	35%	24	60%
News	4.2 (0.22%/95%)	76%	856,913	12%	5.2	57%
Online backup	2.9 (0.15%/4.2%)	10%	7,576	26%	401	-13%
Blogging	0.74 (0.039%/97%)	-34%	487,085	-2.1%	1.6	-32%
Web file sharing	0.32 (0.017%/98%)	-27%	10,822	-22%	31	-6.6%

Table 6: Usage by application categories during January 15-22, 2015. See Table 3 for explanations on columns.

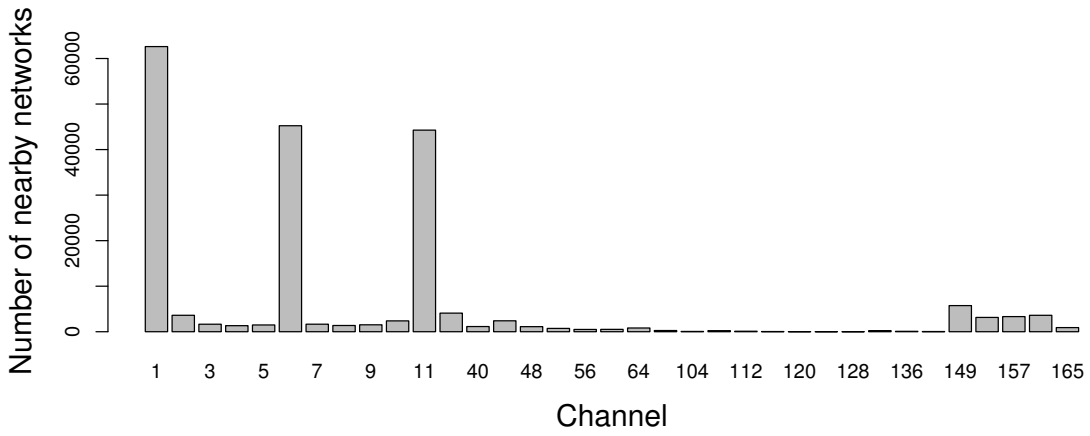

Figure 2: Nearby networks by channel number.

rier sense energy detection, making it more difficult to receive and send packets. Furthermore, some access points are capable of operating as virtual access points, broadcasting multiple SSIDs, which increases channel usage.

Meraki MR16 access points scan for nearby access points whenever clients are not connected and actively transferring data. Figure 2 shows the distribution of detected networks by channel number, demonstrating which channels are most heavily used. This analysis does not include data from any Meraki MR18 access points since they were relatively new during the measurement period and there was not sufficient historical data.

Channels 1-11 reflect the 2.4 GHz spectrum where there are three non-overlapping 20 MHz channels located at channels 1, 6 and 11. Nearby access points are distributed across the non-overlapping channels with channel 1 having around 37% more access points than channels 6 or 11.

In the 5 GHz spectrum, Channels 36 through 48 represent the 5 GHz UNII-1 lower band, channels 52-64 are the UNII-2 middle band, channels 100-140 are the more recently allocated UNII-2 extended band, and channels 149-165 are the UNII-3 upper band. The UNII-2 and UNII-3 bands require the use of a Dynamic Frequency Selection (DFS) protocol where access points first check for the presence of a radar signal and change channels automatically if one exists or is detected during operation. Without DFS bands, there are four non-overlapping 40 MHz channels for 802.11n operation, and with DFS there are ten.

Table 7 compares the average number of nearby networks seen in July 2014 versus January 2015, over the course of a week. In the 2.4 GHz band, the average number of interfering APs (excluding other Meraki devices) is 55, which has grown from 28 six months ago. Given there are only three non-overlapping 2.4 GHz channels, it is likely most Meraki

	Networks	Networks per AP
2.4 GHz (now)	527,087	55.47
2.4 GHz (six months ago)	230,628	28.60
5 GHz (now)	35,010	3.68
5 GHz (six months ago)	19,921	2.47

Table 7: Increase in the number of nearby networks over six months. 9,502 Meraki access points are reporting in January 2015, versus 8,062 access points six months ago. This excludes the SSIDs of nearby Meraki access points.

access point will experience some amount of interference from nearby APs. In the 5 GHz band the average number of non-Meraki access points is 3.68, up from 2.47 six months ago, which suggests it is possible to find a non-overlapping channel, even with 40 MHz channels. We identified that approximately 20% (102,344) of the current nearby networks in the 2.4 GHz band are personal mobile hotspots (Novatel, Pantech, Sierra Wireless, etc.), compared to 56,293 six months ago. In the 5 GHz, only 1.7% of the nearby access networks are mobile hotspots.

4.2 Impact on packet reception

In order to build a table of link metrics for use with mesh routing, each Meraki access point periodically broadcasts a 60-byte packet at 1 megabit/sec on its 2.4 GHz radio and 6 megabits/sec on its 5 GHz radio. These broadcast packets are sent once every 15 seconds and are measured over a sliding window of 300 seconds, and the measurements are recorded by the Meraki backend system.

Figure 3 shows a distribution of link delivery ratios, for 16,583 2.4 GHz links and 5,650 5 GHz links which were reported both six months ago and today. For 2.4 GHz band, the majority of links only receive a fraction of broadcasts, and the overall link delivery ratios have decreased over the past six months. The intermediate links also show variations in delivery rate over time, as demonstrated in Figure 4. In the 5 GHz bands, there are far fewer intermediate links, with over half of the links receiving all of the broadcasts. The 5 GHz links also vary over time as illustrated in Figure 5 but they are more consistent than the 2.4 GHz links.

The data suggests performance as a whole is degrading in the 2.4 GHz band, even for small 60-byte packets. Larger 1,500-byte payloads and frame aggregates will be even more adversely affected, which will have an affect of application performance. Sub-frame error detection and correction techniques such as [16] may be necessary to make effective use of the 2.4 GHz band.

4.3 Impact on packet transmission

As the number of nearby access points grows, each channel is likely to become more heavily utilized, which would reduce the amount of time available for transmissions. Figure 6 shows the fraction of time the carrier sense mechanism was triggered during the last scan interval for each radio, as

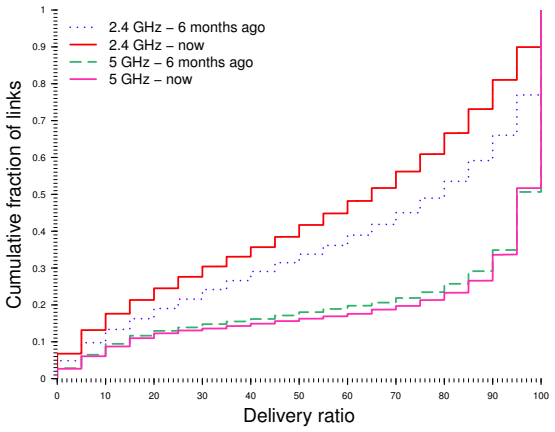

Figure 3: The distribution of the link delivery ratios, six months ago and today. Each data point represents the delivery rate of periodic broadcast packets from one Meraki access point to another where they occupied the same channel.

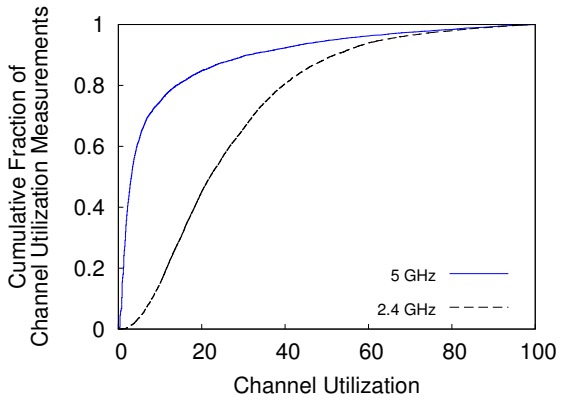

Figure 6: The distribution of channel utilization today in the 2.4 and 5 GHz bands as observed by the Meraki MR16 access points.

measured by the Meraki MR16 access points.

The 2.4 GHz band shows a high level of utilization, with the median access point reporting the energy-detect trigger 25% of the time, and the 90th percentile access point reporting 50%. In the 5 GHz band, channel utilization is lower, with the median access point reporting 5% and the 90th percentile access point reporting 30%. The Meraki MR16 radios can only measure utilization on their current channel, so they do not each provide a view of the complete spectrum. Section 5 examines utilization using the Meraki MR18 radios, which have a dedicated scanning radio that measures utilization across all of the channels.

5. SPECTRUM SCANNING RESULTS

In this section, we take a more complete look at utilization across multiple channels for a given access point. We achieve this through the use of the more recent MR18 access

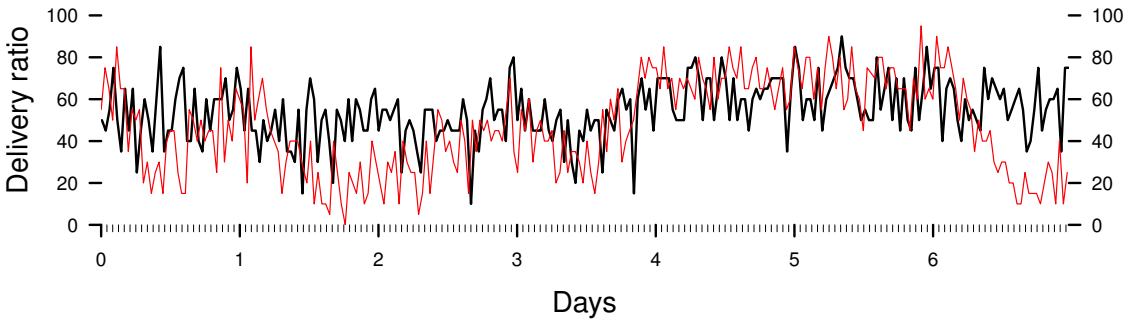

Figure 4: Delivery ratio variation over a week for two randomly chosen 2.4 GHz links.

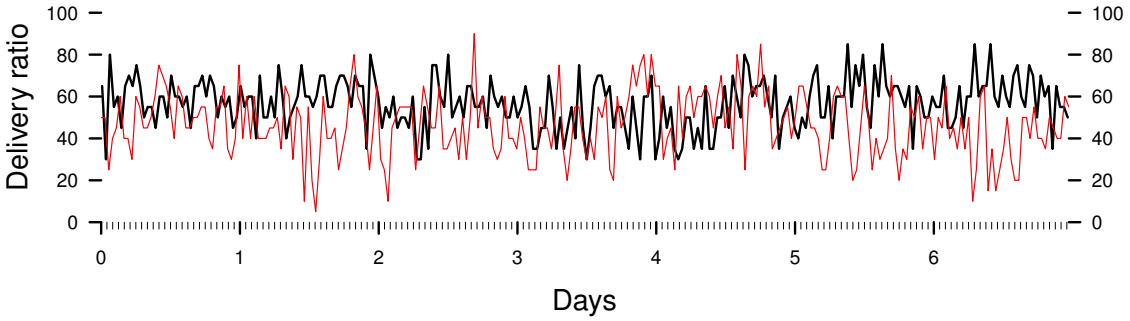

Figure 5: Delivery ratio variation over a week for two randomly chosen 5 GHz links.

point, which includes a third 802.11n radio that is dedicated to scanning the entire 2.4 GHz and 5 GHz spectrum and does not serve clients (unlike the MR16, which only provides utilization on its current channels).

In this section we gather data from 10,000 Meraki MR18 access points, located in the exclusively in the US to simplify complications due to regulatory domains. As it scans, the dedicated radio spends 5 ms on each channel. The backend system collects these results every three minutes, and the results are aggregated over three-minute periods.

5.1 Interfering APs

Figures 7 and 8 plot, for all the access points, the number of access points detected versus the channel utilization for all channels. Because these measurements are taken over three-minute periods (unlike the one-week windows used in Figure 2), these figures provide a more instantaneous view of channel conditions. From the data we do not see a clear correlation between utilization and the number of interferers in either band. This lack of correlation implies that simply using the number of nearby APs is not enough information to accurately select the most available channel and instead it is better to use direct channel utilization measurements.

5.2 Day/night variations

We next examine the relative impact of client usage on channel utilization, by studying variations between day and night. Figure 9 plots CDF of both frequency bands with uti-

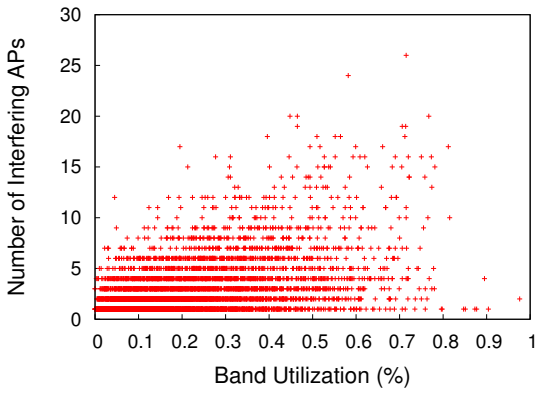

Figure 7: Scatter plot of utilization versus number of nearby APs in the 2.4 GHz band.

lization measurements taken at 10 a.m. and 10 p.m. Pacific time. For the 2.4 GHz band, the median channel (measured at a specific access point), observes around 5% higher utilization during the day versus the night, whereas in 5 GHz the utilization measures are similar.

These results differ from Figure 6 because the Meraki MR16 is only able to measure utilization on its current channel, versus the Meraki MR18 is able to take measurements from all channels. In the 5 GHz band, the majority of channels are unused as seen in Figure 2, which skews the distri-

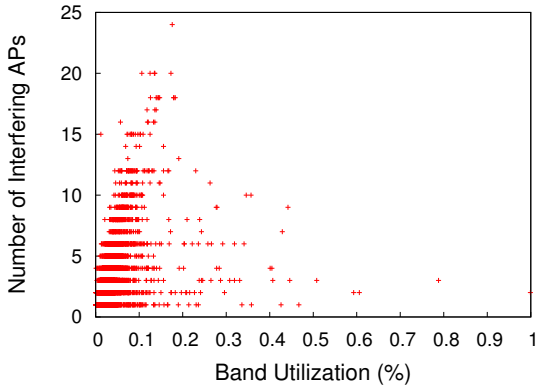

Figure 8: Scatter plot of utilization versus number of nearby APs in the 5 GHz band.

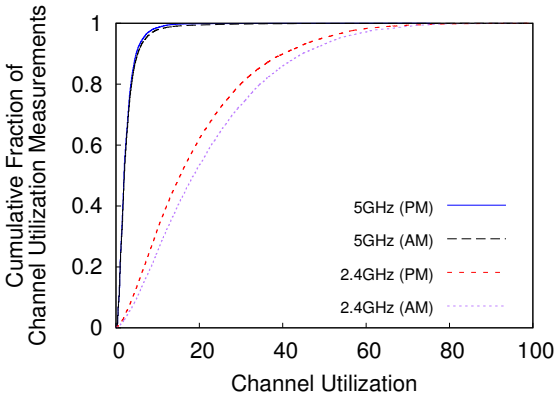

Figure 9: The distribution of channel utilization, during night and day time as observed by the Meraki MR18 access points.

bution towards zero. This also indicates that having a radio dedicated for channel measurement should help in making better channel assignment decisions.

5.3 Presence of Non-802.11 Traffic

For a closer examination of what is using the unlicensed bands, we measure the percentage of utilization that contained decodable 802.11 headers. Similar to the mechanism used to measure the amount of time spent performing energy detection, Atheros chipset exposes a set of microsecond counters which count time receiving 802.11 traffic. These counters look for packets that contain an intact 802.11 PLCP header and preamble (which are sent at a more robust modulation than the data payload) and combine the duration and interframe times to account for 802.11 protocol traffic. These counters do not capture 802.11 frames with corrupt preambles or non-802.11 traffic, which are contained by the overall utilization counter.

Figure 10 shows the majority of the total channel utilization contained decodable 802.11 headers. Furthermore, we inspected the traffic near one access point with a USRP B200 software radio acting as a spectrum analyzer as shown in

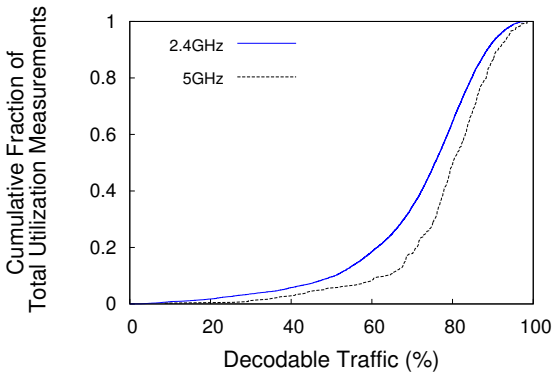

Figure 10: The distribution of decodable traffic (%) in 2.4 and 5 GHz channels.

Figure 11 to examine the structure of the interference. While this view a single instance and is not representative of all access points, we find the 2.4 GHz band to have 22% utilization and includes transmissions from frequency hopping narrow-band interference sources alongside 802.11 frames, similar to observations noted in [11]. The 5 GHz band shows lower average utilization of 2% and is mostly 802.11 transmissions but displays signs of frequency-selective fading as observed in [13].

6. REAL-WORLD EXPERIENCES

The Meraki system was first deployed in 2006, with two models of 802.11b/g radio hardware, for indoor and outdoor environments. Over the next 8 years the system added additional 8 types of wireless access points, ranging from 2x2 and 3x3 802.11n radios, to 802.11ac devices, in both indoor and outdoor variants. The same platform was also used to deploy and manage Ethernet switching equipment and security appliances, which gave the designers a unique understanding of a wide range of network environments.

In this paper, we focused on usage data from two similar indoor radio models to minimize the number of variables that may impact radio link performance measurements. However, building and operating a large-scale system over several years left us with several observations which may be relevant to future work.

6.1 Dealing with bugs in a large system

Debugging large production systems can be complex, particularly in wireless environments where there are a wide range of client devices and physical conditions. The codebase for the `hostapd` portion of the wireless driver, which handles authentication and control traffic for the 802.11 stack is over 350,000 lines of C code.

In addition to the usual practices of using several test harnesses and beta testing, the Meraki system uses a large backend database system to collect information about crashes (firmware and program counter state), along with periodic telemetry about each device's performance, to make it easier to debug problems in the real world.

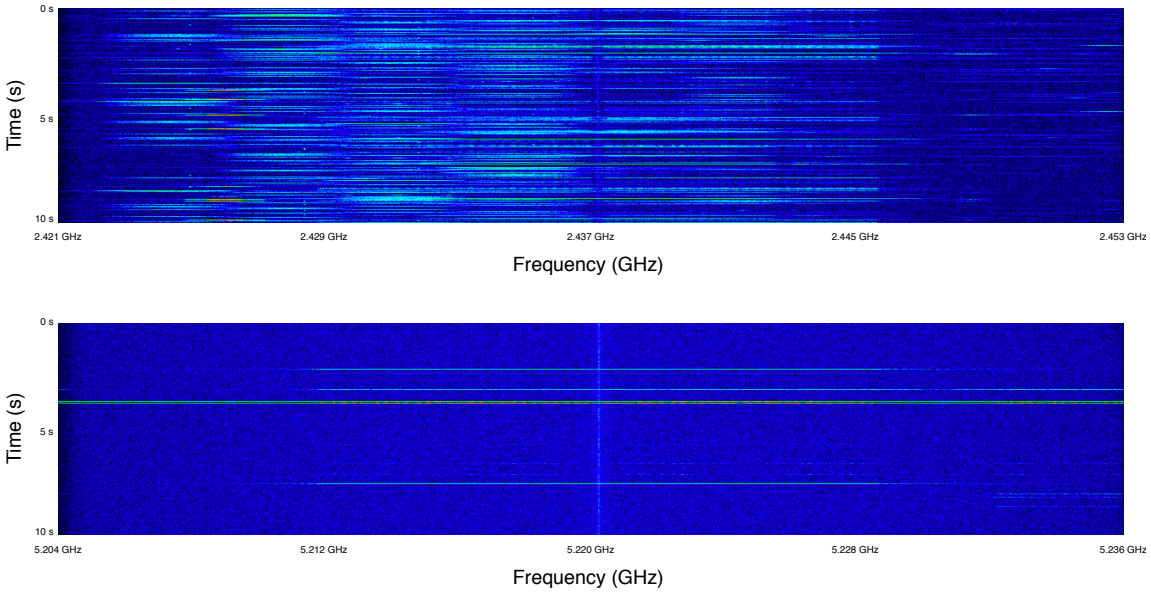

Figure 11: Example spectrum analysis centered at 2.437 GHz and 5.220 GHz, using a 32 MHz wide scan with 4096-point FFT on a USRP B200. The 2.4 GHz scan shows the presence of 20 MHz 802.11 packets, frequency hopping bluetooth with 1 MHz transmissions and other unidentified sources. The 5 GHz scan shows 20 MHz and 40 MHz 802.11 packets and fainter transmissions with frequency selective fading.

As an example of a bug which only surfaced due to the wide range of deployment environments, we received reports of a small number of access points rebooting either minutes or hours after booting on a repeated basis. They all served fewer than a few dozen clients with a typical usage profile. These access points eventually rebooted due to an out-of-memory error (not at the same point in the code). Upon further inspection, we noticed the access points were reporting very large numbers of nearby access points: some of the access points were located in skyscrapers in Manhattan and could decode beacons from miles away, and in another case the access point was on a bus traveling between cities.

Because it is difficult to anticipate such bugs, we found it to be important to measure and instrument the system at large scale and make it possible to examine the system under operation whenever possible.

6.2 Difficulty in predicting user behavior

We also observed client device usage varied over time, which made it difficult for network operators to anticipate usage spikes. In most networks usage between clients was uneven as expected, with a subset of clients driving most of the usage. However, software updates from Apple and Microsoft would drive large downloads across large numbers of clients, sometimes causing sudden increases totaling tens or hundreds of gigabytes.

On a larger timescale, we observed the introduction of smartphones starting around 2006. While we found smartphones consume only around a fifth as much data as laptops, as described in Section 3.2, they exhibit different network connection behavior: roaming across access points, attempting to use cached IP address assignment upon waking from sleep, and implementing aggressive versions of power save poll which increased the data buffered by access points.

Similar to the long-term shift in device types, we observed sustained growth in online video services such as YouTube and Netflix over several years, which increased overall data consumption. While in the past it became common to block heavy bandwidth consuming applications like BitTorrent, we found administrators reluctant to block services such as YouTube given their popularity and wide range of uses.

6.3 Non-wireless problems

Finally, we observed several common problems on networks which resulted in poor performance but were not specific to wireless:

- Overloaded RADIUS/Active Directory servers, resulting in authentication timeouts
- Misconfigured VLANs in campus-scale deployments, either due to numbering or reachability
- Cable problems due to aging building infrastructure, which would result in intermittent connectivity
- MTU configuration and discovery issues
- Upstream bandwidth bottlenecks
- DNS resolution problems
- Protocols like multicast DNS, which work in home environments but cause broadcast issues at campus scale

7. RELATED WORK

Several papers have studied link delivery rates: Aguayo et al. [2] presented data from a 38-node outdoor 802.11b wireless mesh network, and found that the majority of links deliver only a fraction of sent packets and that it is difficult to predict link characteristics from distance or RSSI alone. Reis et al. [18] found similar results with a 15-node indoor 802.11a/b/g testbed. This paper studies a much larger set of approximately 20,000 nodes, with newer 802.11n radios, and finds that even with newer radio technology most links only receive a fraction of packets.

Halperin et. al. [13] studied link delivery in 802.11n and suggested frequency selective fading caused by multipath interference is the primary cause of intermediate links. In their paper, they develop a model for packet delivery based on Channel State Information (CSI) from their Intel radios, which is a better predictor than RSSI alone. The Meraki radios use a different Atheros-based chipset which does not expose CSI in the current driver, but recent work [20, 4] suggests the capability exists for Atheros and Intel radios.

Many papers also indicate non-802.11 sources of interference lead to packet loss. Gummadi et al. [11] showed even weak or narrowband signals from cordless phones, Bluetooth or Zigbee devices can disrupt packet reception or delay transmissions by triggering carrier sense. Gollakota et al. [8] showed similar results with an indoor testbed and studied the physical layer impact of interference at a more granular level before demonstrating software radio-based methods of canceling the interference. In this paper, we quantify the effects of non-802.11 interference on a larger scale but do not employ any active measures for noise cancellation.

Interference alignment and cancellation [10] has also been proposed in different contexts for increasing the throughput of wireless systems. Halperin et al. [12] showed that receivers can recover from colliding packets by employing interference cancellation, and thereby significantly reducing the packet loss rate. Gollakota et al. [9] use interference cancellation to combat hidden terminals, while Lin et al. [15] built a system that enables more nodes to transmit in parallel without harming the ongoing transmissions. Given the observed interference, we expect these mechanisms to improve throughput; however, we currently do not employ any of these techniques in our devices.

There have also been several studies examining usage on 802.11 networks. Rodrig et al. [19] examined usage during a five day technical conference in 2004 and found several link-level phenomena such as bit-rate selection and retransmissions affected overall network capacity.

Ghosh et al. [7] present larger scale data from AT&T's hotspot network, tracing over 243,000 devices over a four-week period. While the data covered several sites, their study was of hotspot users and developed a model of usage distribution in terms of session times and bytes transferred. Their study did not examine specific application usage patterns or present link level measurements.

A 2010 study of Google's WiFi network [1] examined data from 500 outdoor wireless mesh nodes and studied both link-level measurements and usage from 30,000 client devices. Similar to the AT&T study, the Google data primarily focused on outdoor hotspot usage and did not examine the application usage patterns. Both papers also characterized hotspot usage, which differs from campus or office usage.

Gember et al. [6] studied over 32,000 devices connected to the University of Wisconsin's wireless network, with a specific look at application usage inferred from hostname analysis. Similar to this paper, they found web traffic and streaming media to be the largest source of usage, and observed the growing popularity of handheld devices. This paper examines similar types of data but from a much larger set of users and five years later when both applications and devices types have changed.

8. CONCLUSION

Wireless networks using 802.11 have become commonplace, but their performance and link-level characteristics vary because of their use of unlicensed spectrum. In this paper, we conducted a large scale study of several thousand wireless networks, to better understand their behavior.

We surveyed over 20,000 networks and found a wide variety of client devices co-exist, from 2.4 GHz-only 20 MHz mobile devices to three-stream 802.11ac desktop computers supporting 80 MHz channels. Overall client usage grew considerably, driven by bandwidth-intensive applications such as video streaming from YouTube and Netflix and downloading photographs via social media.

We then used measurement data from regular transmissions between access points to study several thousand wireless links over a long period of time. We found that intermediate packet delivery ratios are very common and found in the majority of 2.4 GHz links and many of the 5 GHz links as well. We also noticed the overall delivery ratios of 2.4 GHz links have degraded over the past six months.

Finally, we measured channel utilization using dedicated radio hardware. While most access points share channels with dozens of other access points, but there was no direct relationship between the number of nearby access points and channel utilization. We did, however, observe higher utilization levels during the day, and were able to characterize most of the interference as 802.11, which suggests the 802.11 MAC can avoid interference on the same channel.

These measurements are intended to provide insight into the behavior of unlicensed spectrum for both network and protocol designers. Some practical implications for networks include the importance of (1) traffic shaping at the wireless access point to better serve the growing number of bandwidth hungry clients and applications, and (2) channel planning using a utilization measure to identify the best wireless channel. For protocol designers, the measurements suggest that it is rare to find large blocks of unused unlicensed spectrum and that new protocols will have to co-exist with large amounts of 802.11 traffic.

A copy of the wireless link measurements, nearby networks, and channel utilization data used in this paper is available at http://dl.meraki.net/sigcomm-2015.

9. ACKNOWLEDGEMENTS

We thank our shepherd, Kyle Jamieson, and the anonymous reviewers for their suggestions and feedback. We also thank the engineers who helped build and maintain the Meraki system over the past several years, which made the collection and analysis of this data possible.

10. REFERENCES

[1] M. Afanasyev, T. Chen, G. M. Voelker, and A. C. Snoeren. Usage patterns in an urban wifi network. *Networking, IEEE/ACM Transactions on*, 18(5):1359–1372, 2010.

[2] D. Aguayo, J. Bicket, S. Biswas, G. Judd, and R. Morris. Link-level measurements from an 802.11 b mesh network. In *ACM SIGCOMM Computer Communication Review*, volume 34, pages 121–132. ACM, 2004.

[3] W. Alliance. Total wifi device shipments to surpass ten billion this month, Jan. 2015.

[4] A. Bhartia, Y.-C. Chen, S. Rallapalli, and L. Qiu. Harnessing frequency diversity in wi-fi networks. In *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking*, MobiCom '11, pages 253–264, 2011.

[5] J. Franklin, D. McCoy, P. Tabriz, V. Neagoe, J. V. Randwyk, and D. Sicker. Passive data link layer 802.11 wireless device driver fingerprinting. In *Usenix Security*, 2006.

[6] A. Gember, A. Anand, and A. Akella. A comparative study of handheld and non-handheld traffic in campus wi-fi networks. In *Passive and Active Measurement*, pages 173–183. Springer, 2011.

[7] A. Ghosh, R. Jana, V. Ramaswami, J. Rowland, and N. Shankaranarayanan. Modeling and characterization of large-scale wi-fi traffic in public hot-spots. In *INFOCOM, 2011 Proceedings IEEE*, pages 2921–2929. IEEE, 2011.

[8] S. Gollakota, F. Adib, D. Katabi, and S. Seshan. Clearing the rf smog: making 802.11 n robust to cross-technology interference. In *ACM SIGCOMM Computer Communication Review*, volume 41, pages 170–181. ACM, 2011.

[9] S. Gollakota and D. Katabi. Zigzag decoding: Combating hidden terminals in wireless networks. In *Proceedings of the ACM SIGCOMM 2008 Conference on Data Communication*, SIGCOMM '08, pages 159–170, 2008.

[10] S. Gollakota, S. D. Perli, and D. Katabi. Interference alignment and cancellation. In *Proceedings of the ACM SIGCOMM 2009 Conference on Data Communication*, SIGCOMM '09, pages 159–170, 2009.

[11] R. Gummadi, D. Wetherall, B. Greenstein, and S. Seshan. Understanding and mitigating the impact of rf interference on 802.11 networks. *ACM SIGCOMM Computer Communication Review*, 37(4):385–396, 2007.

[12] D. Halperin, T. Anderson, and D. Wetherall. Taking the sting out of carrier sense: Interference cancellation for wireless lans. In *Proceedings of the 14th ACM International Conference on Mobile Computing and Networking*, MobiCom '08, pages 339–350, 2008.

[13] D. Halperin, W. Hu, A. Sheth, and D. Wetherall. Predictable 802.11 packet delivery from wireless channel measurements. *ACM SIGCOMM Computer Communication Review*, 41(4):159–170, 2011.

[14] E. Kohler, R. Morris, B. Chen, J. Jannotti, and M. F. Kaashoek. The click modular router. *ACM Transactions on Computer Systems (TOCS)*, 18(3):263–297, 2000.

[15] K. C.-J. Lin, S. Gollakota, and D. Katabi. Random access heterogeneous mimo networks. In *Proceedings of the ACM SIGCOMM 2011 Conference*, SIGCOMM '11, pages 146–157, 2011.

[16] A. K. Miu, H. Balakrishnan, and C. E. Koksal. Improving Loss Resilience with Multi-Radio Diversity in Wireless Networks. In *11th ACM MOBICOM Conference*, Cologne, Germany, September 2005.

[17] J. Pang, B. Greenstein, R. Gummadi, S. Seshan, and D. Wetherall. 802.11 user fingerprinting. In *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, pages 99–110. ACM, 2007.

[18] C. Reis, R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan. Measurement-based models of delivery and interference in static wireless networks. *ACM SIGCOMM Computer Communication Review*, 36(4):51–62, 2006.

[19] M. Rodrig, C. Reis, R. Mahajan, D. Wetherall, and J. Zahorjan. Measurement-based characterization of 802.11 in a hotspot setting. In *Proceedings of the 2005 ACM SIGCOMM workshop on Experimental approaches to wireless network design and analysis*, pages 5–10. ACM, 2005.

[20] S. Sen, B. Radunovic, J. Lee, and K.-H. Kim. Cspy: finding the best quality channel without probing. In *Proceedings of the 19th annual international conference on Mobile computing & networking*, pages 267–278. ACM, 2013.

[21] K. Varda. Protocol buffers: Google data interchange format, 2008.