

Roles and Impact on Mobility Network Traffic in Big Data

Jeffrey L. Garner
Indiana University
Online Student
jeffgarn@iu.edu

KEYWORDS

i523, hib315, Big Data, Mobility Network Traffic, Network Forecasting, Data Types

1 INTRODUCTION

At the core of Big Data is a challenge. A challenge of exploration "of the complexities inherently trapped in data, business, and problem-solving systems". [1] Which is by definition, "Big Data".

Imagine a world where your business decisions relate to data sources that range from a flat file from a third-party vendor to millions of internal data records every day, nearly every hour. Add to this, some data sources might "round up" the data, while others relate the data (traffic) to a different geographic standard than others. So it is in the world of mobility network traffic.

2 DATA TYPES AND CHALLENGES

Mobility network traffic providers generate CDR (Call Detail Records) every time a device establishes a connection. These CDRs that are produced by the network equipment provide details about the connection - cell site locations, length of call and device information, including the duration of the call along with other information. It is from these records that the network providers gather, *clean* if need be, consolidate and extrapolate the needed information to bill the customer.

In terms of the CDR data, a large telecommunication provider will create millions of these records every day, even every hour. For companies that have over 50 million devices to manage, and each device can create dozens of records each day, the numbers of records and the size of the data is tremendous. However, with all this data, the management of the records by sheer quantity can lead to qualitative challenges. For example, by the time the millions of records are consolidated to generate files that are more manageable, data details can be lost. While CDRs tell us a great deal, there is much that they do not tell a provider. Therefore, other data sources are used, like data from the network, which provides precise traffic metrics.

This additional network data does not come from the creation of CDRs but is rather collected from the numerous sectors within a mobility network. A sector is a collection of cellular towers and these sectors are in turn gathered together and feed metrics into what are referred to as data collectors. Therefore, the collectors are related to the network vendors that build the equipment. As a result, if the network has more than one equipment vendor, a challenge is to make sure the vendors measure or collect traffic consistently across the network. Once we are insured of consistent measurement of the data from the collectors, we can then consistently map the data into agreed upon geographical areas, known as sub-markets or markets.

This additional network data is free of the challenges and limitations of the CDR based data, this data however, only provides simple traffic measurements. For example, we now know the voice traffic measured in minutes, or the megabyte (MB) traffic in California or South Dakota. But it doesn't tell us the device type or any customer specific data like the CDR data does.

Adding to the challenge, companies like Verizon and AT&T are changing to unlimited plans - which allows the customer complete data freedom, offering package deals with video services and even offering free traffic based on cell phone apps (HBO for free on your device) - So gathering meaningful data on this type of traffic, requires a data set that is much different than simply looking at network or CDR traffic. That is, we need to look at the bits and bytes. We need a much deeper dive into the traffic to start to pull more specific information. For example, we can look at the data packet headers and leveraging an involved process can start to glean an understanding about the network traffic that provides us details and specifics around this big data. For example, we can get data regarding how much traffic is video (directly in relationship to promotions like free HBO), or how much traffic was browsing the web, instant messaging, photo files, VoIP (Voice of IP) and many others.

Additionally, the customer landscape has changed which makes traditional analysis more challenging. For example, in years past, most of the mobility subscribers were *post-paid*. That is, they paid after the actual activity took place. Most mobility subscribers used their mobile device last month and then received their bill this month. Today we have pre-paid customers, wholesale customers and even customers that simply monitor their packages, dog-collars, vending machines and track delivery trucks. We call this the Internet-Of-Things (IoT).

With IoT, lots of projections abound and here is one: "roughly 23 billion active IoT devices by the year 2019 and spending on enterprise IoT products and services will reach \$255 billion globally by 2019, up from \$46.2 billion this year." [3]

Also network providers have learned that nothing puts more traffic on the network like video. Video based apps, like Facebook and YouTube directly impact network traffic. [2]

The impact of apps on the mobility network is significant with no end in site: "when it comes to reaching consumers in mass, the market has confirmed what we have known all along - that we are all building and investing into a platform that can reach heights we may have never seen before. That, to me, is "The WhatsApp Effect", and there is no turning back now." [4]

As shown in Figure 1. You can see the projected video usage increase, by a percentage of the total network traffic over the next five years.

[Figure 1 about here.]

This leaves us with yet another type of data impacting the mobility network that is neither network traffic or data around the traffic types. This is data from the applications. Most of the data from a mobile device is tied to one of many "apps". While the process is wrought with challenges, larger network providers will invest in diving yet deeper into the network and traffic types in order to have a better understanding around traffic specific to apps. Prudence would dictate to a network provider that it is best to know what is on their network. However, strict legal and customer privacy laws, along with application vendors working independently thus not coordinating with network providers, leads to numerous data challenges. While far from a perfect process, gathering as much app level data is critical to the management of any mobility network. Not only to the management of the mobility network but it provides value added information that could impact marketing plans, finance and organizations like strategic planning and technology planning. All in an effort to understand and provide outstanding customer service.

In addition to the efforts of the mobility network provider in gathering data around apps, there are other related data options. Like the saying, "there is an app for that", there are other means of gathering such data. App Annie is one of many companies that provide data related to apps. Both in terms of the numbers of uploads of an app as well as gathering high level app related metrics. As an app developer, imagine knowing the amount of uploads of your app, geographical upload metrics as well as revenue related to the uploads.

Similarly other companies gather app related traffic as a result of their own app that manages customers data packages so that the customers do not use too much data. There are also companies that inform the customer of their intention to gather data based on their usage. This is usually done in such a way that there is no one customers' data that is identified but data from many customers that is combined to provided analytics. Still others have apps that manage the efficiency in not draining too much of the device's battery.

These app level data sources can be critical when a network provider is trying to identify traffic that is no longer on the mobility (cellular) network but has moved to Wi-Fi. Once the traffic is off the mobility network you no longer have data regarding it. All the traditional network data sources are of little, to no, benefit. This importance is magnified when looking at particular apps that can add significant data traffic to the mobility network. For example, Netflix is a heavy Wi-Fi leveraged app but imagine if a percentage of the traffic rolled to the mobility network. So keeping a close eye on the traditional video streaming apps, and it's percentage of usage on Wi-Fi, is a wise decision. As a result, the additional sources can prove critical in building knowledge around your data.

3 CHALLENGE AND CONCLUSION

For mobility network providers, what is the Big Data challenge here? What is the missing piece to the providers that Big Data has an opportunity to help with, if not answer? Providers already have access to network traffic data, along with data around traffic types which is above the OSI Model Network Layer (Open Systems Interconnection) to provide some insights into traffic types; web

browsing traffic, VoIP, video, and even some data around traffic related to apps. The challenge for Big Data is to take all of this data and give network providers accurate analytics on - *customer behavior!*

Can it be done? I believe, with the use of data holistically and with data-driven discovery, it can. However, it is important to note that in order for this to be successful, you have to have a solid understanding of the data itself. It requires an intimate knowledge of the data, the sources, and any underlying limitations and collection challenges. Additionally, it is critical to have substantive data storage capabilities, like data lakes.

A holistic view of the data is to include all the data sources; network data, traffic type data, app level data interrelated and connected hierarchically, so that when you see a jump in the network traffic, you trace the traffic type and app level, which can then lead to accurate deductions to explain the, aberration, one such as, *The Ice Bucket Challenge*, an innocuous social experiment played out on Facebook that demanded a tremendous amount of network capacity. This comprehensive, holistic approach is the only way to paint an accurate picture of user behavior, taming "Big Data" into a beast that can be interpreted. And as a result, helping understand - customer behavior.

At this point we have built a relationship between the various data sources and have let the data drive the results. It's from this process in which we have gained an important business benefit - wisdom. Wisdom gained from a data-driven discovery that can be applied directly to the mobility network itself. From a Big Data challenge, and given data knowledge, we aligned the data and let the data "tell" us the impacts on the network. This wisdom provides us with one last critical benefit for any mobility network provider - a better bottom line, which as they say, is the bottom line.

ACKNOWLEDGMENTS

Many thanks to Professor Gregor Von Laszewski, the Teaching Assistants and Indiana University. I also want to thank Katie, my understanding wife. Lastly, for my employer AT&T for a commitment to education and giving me 26 years of experience, challenge and opportunity.

REFERENCES

- [1] Longbing Cao. 2017. Data Science: Challenges and Directions. Magazine. (Aug. 2017). [https://doi.org/CommunicationsoftheACM\(Assoc. CommunicationMachinery\)](https://doi.org/CommunicationsoftheACM(Assoc.CommunicationMachinery))
- [2] Cisco. 2017. *Cisco - Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021 White Paper*. techreport. Manufacture. <https://doi.org/Manufacture>
- [3] Jay Schofield. 2015. Big Data Challenges Wireless Networks, CIOs. Webpage. (Feb. 2015). <http://www.systemid.com/learn/big-data-challenges-wireless-networks-cios/>
- [4] Semil Shah. 2014. The WhatsApp Effect. Webpage. (March 2014). <https://doi.org/TechCrunch>

4 BIBTEX ISSUES

Warning-unrecognized DOI value [Communications of the ACM (Assoc. Communication Machinery)]

Warning-unrecognized DOI value [Manufacture]

Warning-unrecognized DOI value [Tech Crunch]

(There were 3 warnings)

5 ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

5.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

5.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

5.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

5.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use & but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

`\section{Introduction}` and NOT `\section{INTRODUCTION}`

5.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

5.6 Latex Errors

Erroneous use of quotation marks, i.e. use "quotes" , instead of " "

To emphasize a word, use *emphasize* and not "quote"

When using the characters & # % _ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

5.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inflate your paper if you are below the page limit

5.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named "images"

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use `textwidth` as a parameter for `includegraphics`

Figures should be reasonably sized and often you just need to add `columnwidth`

e.g.

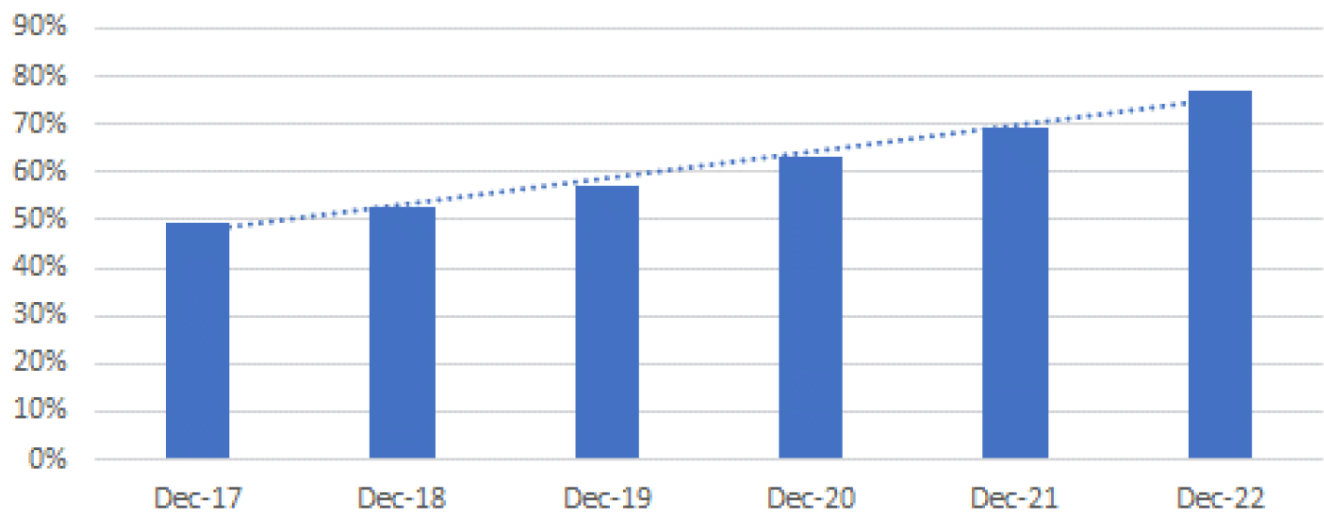
```
/includegraphics[width=\columnwidth]{images/myimage.pdf}
```

LIST OF FIGURES

1 Cisco

6

Percentage Mobile Video Growth Rate



February 2017, Cisco VNI (Visual Network Index) Global Fixed and Mobile Internet Traffic Forecast. Based on video growth at CAGR of 40%.

Figure 1: Cisco