The Intersection of Big Data and IoT

Peter Russell
Indiana University
petrusse@iu.edu

ABSTRACT

Big Data and IoT share a symbiotic relationship with one another that is leading to incredible innovations that were inconceivable just 18 years ago. As a result of this relationship, it has become easier than ever for individuals to customize and monitor various elements of their life if they choose to do so. A project is undertaken to demonstrate how accessible IoT has become to leverage Big Data analysis, how IoT and Big Data are being utilized together in some of the most interesting current use cases and how the technology will need to adapt in coming years.

KEYWORDS

i523, HID 334, Edge Computing, Raspberry Pi, IoT

1 INTRODUCTION

In 2020, it is estimated that 95 percent of electronics will contain IoT technology [41]. This technology, commonly abbreviated for the "Internet of Things", is expected to be so pervasive due to how the technology is defined and how impactful it has already become.

At the highest level, IoT is intended to describe devices that collect and relay information via the Internet. This leaves IoT is broadly defined in the type of application, which could come in the form of a phone, vehicle, a home device like a thermostat or television, but the technology is rather specific in its intended application. That is, these devices are generally made to serve a single purpose and they are extremely adept at that function. In its most powerful applications, massive data sets are created from these individual IoT devices as they are synthesized together to meet a larger need, as we will see.

The specific purpose of an IoT device is what differentiates this emerging technology from traditional computers. With the exception of recent developments, which will be explored in depth, early IoT devices were not intended to do the heavy computing like a computer would do. However, with rapid advancements in computing power and speed, the line differentiating the two has begun to blur. It is this increase in computing power that has lead IoT and Big Data to have a cooperative relationship to create some of the most exciting technology that's available today.

The ability to collect and process more data has increased the utility of these devices as they're able to become more personalized, spurring tremendous growth recently, on the order of 30% annually. In 2017, it is expected to be the year that the number of IoT devices exceeds the number of people on Earth [18]. This personalization is not without consequence though, which will be discussed later, so the relationship between IoT and Big Data is still evolving.

To begin, we examine how the IoT came to be and continuously evaluate how it is integrated with Big Data. Then, a demonstrative project will be outlined to show the Big Data can be used in an IoT device. Next, we discuss high level implementations of these technologies in modern use before discussing some of the challenges the industry is facing.

2 EMERGENCE OF IOT

Given the massive and recent popularity of IoT, it might be a surprise to some to learn that this concept has been around since 1999. The idea to have multiple, remote devices communicating with one another to gain insights to a single problems was originally conceived by Kevin Ashton as a solution to supply chain management [16]. At that point, the idea was ahead of its time as the internet was still gaining widespread adoption. However, as computing power and sensor costs have declined, IoT has become a main an indispensable aspect of most people's lives. One such example could be the integration of global positioning systems (GPS) into cellphones, which was introduced in just 2004, but has become a staple for nearly every phone released [50].

2.1 What Defines IoT?

Given the ascension of so many new technologies, it could be helpful to understand what technically constitutes an IoT device. This will be useful later when discussing the sample project undertaken and how these both relate to Big Data analysis.

As stated earlier, at a high level, IoT is meant to describe devices that use internal or external sensors to connect to the Internet. These sensors could come in the form of the well known types, such as Wi-Fi, Bluetooth or RFID, to the perhaps less widely known, such as NFC (Near Field Communication) or Zigbee [43].

For these IoT devices, the Internet allows them to be tremendously influential in the advancement of Big Data by virtual of the amount of data the devices are able to collect. Specifically, IoT allows its users to quantify the world around them in ways that were previously not possible. For corporations, this yields tremendous advantages when it comes to business planning or equipment monitoring. For example, the average wind farm can generate 150,000 data points *per second* and an engine turbine could give 500 gigabytes of data [28]. Additionally, for individuals, IoT enables people to monitor their activity on a daily basis through wearable fitness devices and customize their homes to save on energy consumption. It has been estimated the average household generates approximately 2,000 gigabytes of data a year and this is expected to increase five fold by 2020 [52]. As we will explore, this rapid increase is due in large part to the computing power of the individual devices, which allow for a greater volume of data to be collected. For example, if a person enjoys a simple bike ride and purchases the Garmin Edge 500 watch, on a single ride they are producing data across 61 different variables for statistics such as heart rate, elevation gained, cadence and output produced continuously for the duration of their ride [17].

3 IOT PROJECT

3.1 IoT Device

The Raspberry Pi 3 (Model B) was chosen as an example IoT device to demonstrate how these devices can be used to from Big Data analytics. The Raspberry Pi has drawn tremendous accolades for its initiative to get inexpensive, but powerful computing power into the hands of aspiring programmers and hobbyists. Equipped with 1GB of RAM, a 1.2GHz quad-core processor and Bluetooth/Wi-Fi capability, one can purchase the device for just $35.

3.2 Description

The goal of this device is to create a personalized interface that gives the user a morning snapshot for relevant, important information to begin their day. As it relates to IoT, this project uses IoT technology through Wi-Fi to source the output of Big Data projects undertaken by others (ie. Google and Weather Underground as will be shown).

3.3 Implementation

For those unfamiliar with the Raspberry Pi, the initial setup could be somewhat intimidating the first time around. Specifically, the Raspberry Pi comes as a truly blank slate and to begin using it, one will need to write the OS, Raspbian, onto the Pi. Several tutorials are available online to get to the desktop, so in the interest of brevity, the discussion below will assume that the user has been able to successfully get the Pi operational and to the Linux prompt with Python installed to run the script.

The application was developed using Python, utilizing the Kivy package for GUI development, the requests and Beautiful Soup packages for the user location, news stories and sports scores along with Yahoo Weather/Weather Underground via the Weather package. Outside of these, standard Python library packages were used.

3.4 Results

As part of the display, a continuous running clock was added, which necessitated the application to be on a constant refresh. This was implemented successfully and at a relatively low cost with no significant delays. The total build of the application consumed 966,565 bytes with each refresh using just 1032 bytes. At the initialization of the program, the application uses the user's IP address to find their zip code to populate the local weather forecast and local news. For the weather, the user will get the current temperature with a high/low for the current day and each day in the five day forecast.

Additionally, as news stories are published to the WSJ news feed, they will be read into the application and refreshed. Stories are shown in chronological order along with their time stamp of publication and each is hyperlinked directly to the full story if the user wants more information.

3.5 Application to Big Data

One of the main benefits of IoT is synthesizing data across numerous platforms and data sources into a desired output. Our desired output is a one-stop interface with interesting information that can be displayed anywhere with an outlet, internet connection and monitor.

The various components of the monitor are only made possible by the creativity by the providers in solving difficult Big Data issues, such as the clustering of news (Google) and weather forecasting (Yahoo/Weather Underground). Each provider's relationship to Big Data is worth examination and will be the topic of the following sections.

3.6 Google News

3.6.1 Big Data Description. Google News has evolved into a central source of information for how a large share of the population receives its news. In fact, as a display of the trust that users place in Google to deliver the most information in the most efficient manner, it was found that users are more likely to trust a Google news headline than that same headline from the original source [26]. Additionally, 44% of users were found to read nothing more than the headlines [22]. This is a testament to their ability to simplify the universe of world news into succinct rankings.

Entering into its fifteenth year, Google News aggregates from 50,000 news sources worldwide across 30 different languages. In 2012, they reported the division was receiving 1 billion unique visits a *week*[4]. For reference, major individual news providers, such as CNN and the New York Times receive 125 million and 99 million unique visitors per *month* [3]. These statistics further demonstrate Google's successful navigation of the Big Data problem for news stories in the eyes of its users. It's relatively clear why this is an important Big Data problem, but one might be curious how they're able to effectively navigate the problem. Unfortunately, the full design from start to finish is a well kept secret, but pieces have been released and one can piece together a mosaic view of what might be going on under the hood. The decision to not disclose the techniques for ranking news stories is understandable, but it has been a lightening rod of controversy nonetheless. Some view the decision of Google's scoring as effectively acting as a censor for the internet while they maintain it is to keep the integrity of the algorithm so that news stories cannot be written purely for traffic, known as search engine optimization [12].

On the surface, one might question the economic value of Google News to the larger company since it is a free service for both users and providers. However, there is a tremendous amount to be gained in solving this Big Data problem. Even though Google does not show ads on its news site, it was estimated to generate spillover traffic into its search engine that leaves the News entity worth $100 million in 2008 [60]. The current valuation is undoubtedly higher, but it remains undisclosed. So while there are profits to be made for Google in this quest, publishers of these stories have a tremendous amount of interest in this problem as well. Some providers don't believe content should be indexed to Google's search algorithm for free and Google should pay them for their investigative research. One such provider, who happened to be Germany's largest news source, decided to remove themselves from the index for two weeks. The results were devastating for the site as traffic through the site dropped by 40% [59]. It was a quick lesson in how critical positioning can be be in the Big Data world of news aggregation.

3.6.2 Big Data Solution. Just as news is constantly evolving, so are Google's solutions to this Big Data problem. As we've seen recently, news aggregation services are under pressure to become

more intelligent on what news is shown in the hopes of preventing fake news from making it into the top results.

The technical specifics of what Google is implemented has largely been kept under wraps, but we did learn a few of the techniques and platforms in 2007. In at least the early versions, Google used MinHash, Probabilistic Latent Semantic Indexing and Covisitation to solve this Big Data. Specifically, these methods will compare historical clicks with other similar users for recommendations, decipher key words and phrases from an article for grouping and track how news stories are clicked within a certain time frame to find stories that were read successively. For processing these queries, Google uses MapReduce and Hadoop architecture [55].

For the inputs into these algorithms, Google will analyze several metrics of a provider to see how they should be ranked along with the user preferences. These metrics include things like how large the staff is, how many articles they put out, how many websites reference that news source (PageRank) and the breadth of news topics covered [15].

3.7 Weather Underground via Yahoo API

3.7.1 Big Data Description. Weather is a primary concern in business planning for many industries, such as airlines or agriculture. As a result, companies are willing to dedicate a tremendous amount of resources towards accurate forecasts. One of the most innovative companies and a great example of the intersection of IoT and Big Data is Weather Underground.

Weather Underground is a weather forecasting service that was once owned by the Weather Channel and recently, partially by IBM to integrate with its growing IoT ecosystem. What makes the company unique is how their forecasts are formulated. In their model, they couple traditional forecasting tools with IoT. The traditional readings come from the National Weather Service (NWS), which aggregates data from airports and weather balloons. IoT has lead to a new dimension of forecasting as personalized weather stations are distributed to to its users for live forecasts in places that traditional instruments might not be available. As of now, they have 250,000 users set up on the platform. This setup provides an additional layer of information, yielding more frequent data, longer forecast windows and greater certainty for a given area. Namely, users can get new forecasts every 15 minutes (versus every 4 hours on the NWS) and forecasts up to two weeks in advance (compared to one week for NWS) [54]. This use of the IoT, specifically edge computing, which will be expanded on later, provides a tremendous example of how IoT can be used to enhance Big Data analysis.

For those that choose to participate in the service, they will purchase a Personal Weather Station (PWS) that allows them to measure temperature, humidity, pressure, rainfall, wind speed and direction via sensors. The major advantage of the PWS comes from its pressure and wind metrics as users can get a better idea of humidity and wind chill, giving a more accurate representation of current conditions. Neither of these are available through the NWS. In the end, this amounts to around 3 billion data points for the Weather Underground model, servicing around 26 billion inquiries a day [24].

3.7.2 Big Data Solution. To process its data in the past, which amounts to multiple terabytes daily, Weather Underground has stored its forecasts, radar data and satellite data using Apache Hadoop and Amazon Web Services [47]. In fact, IBM has stated a large reason for their motivation to have an ownership stake in the company was due to the cloud infrastructure that Weather Underground had built for fielding the massive volume of requests and forecasts it processes daily.

3.8 Future Considerations

It is rather commonplace knowledge, for better or worse, that the apps we use daily are collecting data on us. This data aggregation is one of the main debates around IoT. Ironically, one of IoT's primary benefits makes it also one of the most unsettling for others, fearing how the data could be used in the wrong hands. In fact, in 2014, it was found that of the top 200 free apps in the Apple store, 95% were engaging in "risky behavior" [36]. These risky behaviors, which we will explore later, are defined as activities such as tracking locations, accessing users' contact lists or selling registration data to ad agencies.

In subsequent builds of this application, if the intention was distribute to a wider audience, collecting volunteered user data would be an interesting addition. Then, this data can be pooled together for how the application could be tailored to meet geographic or demographic preferences.

Additionally, as with most apps, we would be interested in tracking the number of downloads, active users and which panels of the display are clicked most often. All of these metrics can be readily accessed through integration with Google Analytics, which allows one to analyze different events within the application [20].

4 EDGE COMPUTING

With the Personal Weather Systems, we were able to see how IoT can complement Big Data analysis. This is an example of an emerging technology known as "edge computing" that is transforming how the cloud is utilized.

Edge computing gains its name from how the information being processed by the device. Prior to this recent innovation, information was gathered, sent to the cloud, processed there, and then the output is pushed back to the device. Namely, it was a centralized process. However, with edge computing, devices are more intelligent in what information they choose to send, providing a much more efficient process. For example, rather than having a camera monitor an area constantly, even when there is no motion, modern IoT cameras have been equipped with motion detection so information is only sent when there is something to actually record. Since this decision and processing is made on the actual device, it is considered to be at the *edge* of the network.

Traditionally, IoT devices that were intended to work in conjunction, such as surveillance cameras, were simple in their functionality and storage. Namely, a group of cameras would record individually and send their results back to a central server. However, with improvements in image quality, this becomes a Big Data problem very quickly as these cameras are running around the clock collecting footage. In the historical model of a centralized server, this setup eventually creates problems as bandwith and storage issues emerge. These limitations are the problems that edge computing seeks to

circumvent and has become a major catalyst in the growth of IoT devices [51].

Circling back to the original project that was undertaken, the application benefited from edge computing through the weather data, but the device itself serves as a great example of why edge computing is even possible in the first place. Specifically, it is possible due to the dramatic decrease in computing costs. For the cost of $10 one can get a single-board computer with 1 GHz and 512 MB RAM through the Raspberry Pi. This type of processing is close to becoming the majority as it is expected that by 2019, 45% of all data collected by IoT devices will be processed at the edge of the network [37]. As we will see, this technology is allowing early adopters to gain unique, real-time insights through Big Data analytics into the health and composition of their businesses.

4.1 Use Case: Fraud Detection

Fraudulent transactions represent just 1% of all transactions. However, while the relative size of these transactions to the overall market are small, their absolute impact is enormously detrimental to merchants and financial services companies. In 2015, total fraudulent transactions created damages of $22 billion [29].

The economic impact of these transactions has given these companies a tremendous incentive to innovate their way out of this problem. The marriage of IoT and Big Data has now provided them the opportunity to have near real-time analytics, which is necessity to effectively manage the problem. This is because the approval process for a transaction needs to be as close to instantaneous as possible. If shortcuts are taken in the analysis to increase speed, fraudulent transactions could slip through and not get flagged. IoT has helped make this trade-off between accuracy and speed less of an issue with new innovations, such as Visa's Ready program.

Visa Ready is an innovative program enables payments through IoT for both security and convenience. Instead of traditional means of payment authorization, such as simply swiping your credit card at a vendor, IoT enables Visa to take advantage of improvements in biometric technology [57]. Visa has introduced multi-dimension verification through biometrics by letting users endorse a payment through their fingerprint, iris scan, face scan and even their voice [58]. This type of technology is gaining adoption and there are expected for be 500 million devices with biometric sensors by 2018 and 26 billion by 2020 [48].

Complementing biometric data, as IoT devices become more mainstream, companies such as FICO are using behavioral data in fortifying their analysis of whether a transaction is fraudulent or not. This type of analysis is not new in and of itself as it has been established as a way to identify e-commerce fraud, but the application through IoT is providing a new dimension of analysis. Traditionally, behavior data was tracked to see how a user interacts with a website to reduce the number of false positives that get flagged, which could occur if a user was on a business trip and abruptly logged into their account to buy something from an IP in another country [14]. With IoT, this adds a tremendous amount of data to an already Big Data problem. Now these companies will have data on how users interact with an IoT device, such as how they hold their device in the case of a phone or their tendencies when using the keyboard [25]. From a business perspective, this all occurs in the background without the user's experience without the product being interrupted.

As a testament to the future of this relationship between IoT and Big Data, Visa has partnered with IBM. This was done in an effort to gain maximum benefit from this new biometric technology by leveraging Visa's payment infrastructure with IBM's efforts in artificial intelligence and Big Data analysis with IBM Watson [31].

4.2 Use Case: Autonomous Vehicles

In many ways, autonomous vehicles represent the pinnacle of edge computing to date in unifying IoT and Big Data. Among its many goals, this technology is trying to use Big Data to resolve one of the modern tragic realities of our modern world - automobile fatalities. Automobile accidents cause 1.2 million deaths a year, 94% of which are attributable to human error [30]. For this reason, in conjunction with expected energy savings from car designs with this technology, the technology is expected to experience adoption rates that rivals mobile phones with significantly more impact [23]. Traditional car makers have taken notice of the potential future and as an example of this, General Motors recently hired an Uber engineer to lead its self-driving initiative as the company's first ever Chief Technology Officer [5].

The relation of autonomous cars to IoT via edge computing is once again out of necessity for real-time functionality. A car that processes it should stop two seconds too late is as potentially useful as never making the calculation in the first place, so timing is of the utmost importance. Amazing progress has already been made in the speed and complexity of calculations these autonomous vehicles can handle. One of the highest profile graphic card manufacturers, Nvidia, recently announced their system for autonomous vehicles at the rate of 320 *trillion* operations a second [21]. Since these vehicles are equipped with various types of sensors to process its environment, this type of computing power is a near necessity to tackle this Big Data problem in real-time.

Kevin Ashton's original vision for the IoT was to have an accurate view of inventory as RFID scanners synced over the Internet. In just 18 short years, these autonomous vehicles are achieving the same end of communication with one another on an incredible scale. In what's known as "vehicle to vehicle communication" autonomous cars will be able to send one another information on important considerations, such as road hazards or conditions, allowing GPS to take the most optimal route to its destination. Similarly, speed limit signs can take weather conditions into account, dynamically adjust the speed limit of the road and relay this to the car's navigation system [1].

The companies that are pursuing autonomous driving are largely having the cars learn through the experiences of its sensors. It would be impossible to code every possible scenario a car could face, so instead, data is collected from the various sensors and loaded to the cloud for later analysis. For example, Tesla is accumulating a million miles worth of data across its sensors every 10 hours, leaving it with 780 million through mid-2016 [7]. These sensors on board, which will be briefly described to show their application, are expected to generate 4,000 gigabytes of data *daily* [38] [19]. This is another instance of the familiar union between IoT and Big Data.

4.3 Use Case: Health Care

The United States, like the world as a whole, is experiencing an aging crisis in its population. In both the world and the United States, the number of adults aged 65 and over is expected to double by 2025. In the United States, this demographic of the population will move from 15% to 25%. While this jump is not negligible, the most alarming aspect of this statistic is that in 2010, the elderly portion of the population was just 10%, but accounted for 34% of medical expenditures [34].

For this reason of high future expenditures, much of today's public policy debates center around how resources will be pooled to meet this not so distant future need. Currently, one of the most promising use cases for edge computing is coming from health care and how the technology can be used to provide better care to a wider range of people.

Through edge computing, doctors have the ability to gain insights into their patients through sensors that can be worn by their patients, such as a heart monitor. This allows for early identification of irregular patterns and allows for an earlier diagnosis, potentially saving the patient's life compared to earlier times when a heart attack could strike abruptly without warning. This usage is directly related to Big Data as doctors now can get continuous, real-time assessments of their patients. This makes way to more accurate future diagnoses as more insights can be gleaned between the true cause and effect of a particular ailment [42].

Outside of data analysis by doctors, the patients themselves are expected to receive numerous benefits from this type of monitoring. Namely, those who are less mobile no longer need to make a physical trip to see the doctor as the doctor has the diagnostics they typically need and at a much more granular level [6].

In addition to the elderly, this type of real-time feedback system through edge computing can be incredibly transformative for those with health conditions that require nearly continuous monitoring. One such example has been demonstrated with epileptic patients. An edge computing solution has been introduced that epileptic patients can use and if a patient experiences an epileptic episode, an immediate alert is sent to family members and doctors [53]. This type of technology is only possible through edge computing because the alerts are triggered by monitoring historical metrics versus live readings in areas like heart rate and sudden movements. The delay that would be incurred by sending this data to the cloud and waiting for a response would have too much latency to be an effective solution to this problem.

Another promising area for edge computing within health care is for those suffering from mental diseases, such as dementia or Alzheimer's. With this technology, family members can monitor and set alerts if a particular perimeter is breached from where their loved one is supposed to be staying [8].

4.4 Use Case: Retail Shopping

Worth $2.6 trillion, the United States retail industry comprises 15% of national gross domestic product [13]. The ground is shifting underneath this industry though as brick and mortar stores are under siege from a surging market share by Amazon, which is up 150% since 2013.

These traditional stores still hold the top rankings in the retail sales by size, but the ability of Amazon to utilize Big Data for a personalized shopping experience online is forcing these top retailers to adapt with a competing level of customization. Amazon's recommendation engine allows them to see into a user's purchase history, viewing history, rating history and search history, which are all used to point the customer to the most likely product they're looking for. In fact, Amazon is even working on an IoT sensor that they intend will act as a personalized stylist. The device will take a picture of your outfit and make recommendations of what would look best, based on the recommendations of its algorithms that are supplemented by fashion stylists to reflect current trends [33]. As a result, IoT gives Amazon a level of scalability to its entire customer base to create more information and data about the customer that is simply not available to the brick and mortar stores.

To try and compete with this personalization though, brick and mortar retailers are using edge computing to introduce technology that was science fiction 15 years ago in the movie Minority Report. In the movie, which takes place in 2054, the main character is rushing through a busy shopping center when he passes various kiosks that address him by name and ask about his recent purchases in the store. This is the reality that retailers are now using through real-time facial recognition, enabled by edge computing to integrate IoT and Big Data. With this, they are also collecting broader demographic statistics by tracking customers' ages, ethnicity and gender [32]. In fact, America's largest retailer, Wal-Mart, is currently using facial technology to sense customer's moods and find those who are dissatisfied [40].

While we haven't quite hit the personalization depicted in Minority Report for the general public, those with celebrity can expect that high-end stores they visit will recognize them upon entry. For example, one such jewelry store in Los Angeles is equipped with facial recognition technology, stocked with a database of celebrity pictures from Google Images and when someone is recognized, an alert is sent to the manager with purchase history and sizes [46].

Outside of custom shopping, facial recognition is also being used by traditional stores to deal with a risk that e-commerce is not exposed to - shoplifting. With this technology, a retail store can identify when a known shoplifter is most likely to re-visit the store and when, which were previously unquantifiable. Once they are identified on site, management is sent an instant alert and the customer is escorted from the store to prevent further loss in the future [11]. Additionally, RFID sensors are being used on items individually to better track items outside of the store for loss prevention like this and better supply chain management [10].

5 CONCERNS WITH IOT

As exciting as these use cases are about what the future might hold, innovation is outpacing legislation for IoT. As we will expand upon below, a race to release products has left consumers susceptible to hacking in some cases as security measures have not been fully developed yet for these devices. Additionally, with the customization that comes with IoT, consumer information is being sold to advertising agencies in many cases without the consumer's knowledge.

5.1 Security

While we have discussed some of the most exciting and interesting developments in IoT, this blistering pace of innovation has come at a price. There are experts in this field that believe the connectivity of these devices are a gateway of vulnerability as many IoT devices do not have sufficient security measures, allowing malicious actors direct access into some of people's most private details.

For most utilizing IoT, the technology is used to make their lives easier in some respect. However, when it comes to security, it is believed this approach of a "hands off" relationship with IoT leaves users susceptible to security breaches. Specifically, users need to be diligent in making sure their software is up to date across *all* devices. The reason for this is that with a large network of IoT devices, hackers now have multiple fronts on which they get behind the firewall whereas their only avenues traditionally were the computer and more recently, smart phones. As a result, negligence in one area could be enough of an opening for a comprised network where hackers could take control of a device, which is particularly worrisome in the case of an autonomous car.

Another dimension of risk for IoT security sits with the creators of this technology. Underlying in the assumption about users being diligent in updating their software to prevent breaches is that the developers of the software are actually making continuous updates to adapt along with hackers. However, as time goes on, new products are likely to draw a company's limited resources away from maintaining older products.

In response to these risks, two significant changes have been undertaken to mitigate some of the risks. Namely, companies have introduced automatic updates and used the same operating system across later models of a particular product. These automatic updates then take the burden off of the user of IoT technology, which is an attractive feature as many adopt the technology to simplify their daily life. Additionally, when companies are able to use the same underlying operating system across later products, they're able to update all products in lockstep with the developing security community, ensuring no older products are left behind as an opening behind the firewall [39].

Fortunately, these security concerns with IoT have largely played out in the hypothetical. In fact, surveys have found that the majority of consumers are unaware of IoT security risks and once made aware, do not consider the risks serious. In fact, surveyors even found that if a device had a known security flaw, 20% of consumers are still willing to buy the product [9].

For this reason, with no major attacks to date, adopters of IoT have possibly felt insulated as an overwhelming majority are not threatened by the security risks IoT could pose. This is not to insinuate that IoT attacks do not regularly happen, but instead that they have not occurred on the scale that some of the largest security breaches in recent years have occurred, such as the Target Corporation's incident in 2013. In that breach, 110 million consumer credit card numbers were stolen, along with personally identifying information like their address, e-mail and phone number. The entire episode was estimated to have cost Target $162 million [2].

While an IoT originated attack like this has not happened yet on this scale, these attacks do occur with frequency. One such statistic demonstrating this unsettling fact is that half of all companies that have adopted some element of IoT technology have experienced a security breach. In the end, these breaches have cost an average of 13% of annual revenue [45].

The closest demonstration of IoT risks came in October 2016 through the "Mirai" malware, which was used to attack DNS servers and bring down high traffic websites, such as Netflix and Amazon. Disturbingly, "Mirai" translates to "future" in Japanese. With Mirai, the program is continuously scanning the internet for IoT connected devices that have left the default user name and password. Then, once a device is found, it is turned into a bot that is used to amplify a DDoS attack. Incredibly, the average IoT device is scanned every two minutes with this bot, leaving an extremely small margin for error in being compromised [44].

This breach demonstrated the downside of the highly connected nature of IoT. Against the benefit of having devices that can communicate with one another, in the event of an attack, these devices are intertwined and will be equally compromised. The network of IoT devices has gotten so complicated for some companies that one survey found 66% of IT professionals aren't sure how many devices are in their environment [35].

5.2 Privacy

Naturally, one of the consequences of a security breach via an IoT device would be having personal information comprised. However, outside of this direct relationship, there are concerns on privacy as it relates to usage as laws are behind technology in how this data can be used. The only major pieces of legislation that concern privacy at the federal level are through HIPAA for medical records and the Fair Credit and Reporting Act. Outside of these, the task of regulating privacy is left to states, which are behind the curve in today's fast paced, data driven world.

In a similar conundrum as the security concerns with IoT, one of its greatest features in its ability to continuously monitor and collect this data into Big Data sets is also the reason some hold reservations on the technology. This is mainly due to the fact that this data is not collected into a central repository, like your credit, to see what information is being associated with you. To take it a step further, it is not even clear who has what data on a particular user.

In a shock to most on how little personal privacy may exist in our technology saturated world, it was discovered that the CIA and MI-5 intelligence agencies were using "smart" TVs to eavesdrop on conversations in people's homes. For security experts, this was no surprise and known to be an easily accessible device, but those outside of that community felt an invasion of privacy [49]. Discovered in 2016, the program was used in 2014 by exploiting the voice enabled features that Samsung included in its TVs to listen to conversations. The power button was even programmed to look as if the TV was off while this recording was happening [56].

While this spying was alleged to have just been on "people of interest", the average consumer with a smart TV has likely experienced spying they were unaware of through their viewing habits. By default, Vizio TVs were have found to be recording their customers activities by logging metrics such as date, time, show, whether it was live or recorded and how long it was watched. This is estimated to have affected 11 million TVs in the end before the

FTC outlawed the practice of having these settings turned on by default [27]. This would be a utilization of IoT and Big Data that few would be comfortable forfeiting without their consent.

6 CONCLUSION

As we've seen, IoT cannot realize its full potential without Big Data. The IoT universe represents the senses by which Big Data is collected for later insights and innovations. For this reason, the IoT revolution has the potential to completely change the world as we currently know. It could be a world in which automobile accidents are no longer a tragic reality or a world where health care delivers the most personalized plan with attention on every minute detail. Additionally, users are able to benefit from the increase in computing power per dollar spent, allowing them more flexibility than ever to design their own IoT device, as was demonstrated in the application made for this paper. However, against this rapid pace of innovation in IoT, some of its most attractive features of interdependency among devices expose the technology to some of its greatest vulnerabilities. Keeping this growth rate in the products in step with security will prove to be one of the biggest challenges in coming years.

A CODE COMPILATION AND SAMPLE OUTPUT

The following urls are intended to direct to various parts of the project.

- Packages required to compile the project along with sample input
  - https://github.com/bigdata-i523/hid334/tree/master/project
- Python code to create the monitor:
  - https://github.com/bigdata-i523/hid334/blob/master/project/code/project.py
- Weather codes:
  - https://github.com/bigdata-i523/hid334/blob/master/project/weathercodes.py
- Kivy file:
  - https://github.com/bigdata-i523/hid334/blob/master/project/DailyView.kv

ACKNOWLEDGMENTS

The author would like to thank Professor Dr. Gregor von Laszewski, Juliette Zerick and the other Associate Instructors for their support and suggestions in exploring this topic.

REFERENCES

[1] Philip Adams. 2017. Why self-driving cars can't start without edge computing. Website. (07 2017). https://knect365.com/cloud-enterprise-tech/article/b4751c4b-7b5d-4407-8789-420289799988/autonomous-cars-cant-start-without-edge-computing

[2] Taylor Armerding. 2017. The 16 biggest data breaches of the 21st century. (10 2017). https://www.csoonline.com/article/2130877/data-breach/the-16-biggest-data-breaches-of-the-21st-century.html

[3] Jeremy Barr. 2016. The New York Times Pulls Back Ahead of the Washington Post for Unique Visitors. Website. (02 2016). http://adage.com/article/media/york-times-pulls-back-ahead-washington-post/302720/

[4] Krishna Bharat. 2012. Google News turns 10. Website. (09 2012). https://blog.google/topics/journalism-news/google-news-turns-10/

[5] Johana Bhuiyan. 2017. GMfis self-driving division has hired a former top Uber engineer as its first CTO. Website. (11 2017). https://www.recode.net/2017/11/30/16720994/gm-cruise-cto-susan-fowler

[6] Isaac Christiansen. 2017. The Internet of Things and the Evolution of Elderly Care. Website. (06 2017). http://www.iotevolutionworld.com/smart-home/articles/432936-internet-things-the-evolution-elderly-care.htm

[7] Michael Coren. 2016. Tesla has 780 million miles of driving data, and adds another million every 10 hours. Website. (05 2016). https://qz.com/694520/tesla-has-780-million-miles-of-driving-data-and-adds-another-million-every-10-hours/

[8] Reenita Das. 2017. 10 Ways The Internet of Medical Things Is Revolutionizing Senior Care. (05 2017). https://www.forbes.com/sites/reenitadas/2017/05/22/10-ways-internet-of-medical-things-is-revolutionizing-senior-care/#5e01a7965c8f

[9] Gary Davis. 2017. A Cybersecurity Carol: Key Takeaways From This Yearfis Most Hackable Holiday Gifts. Website. (11 2017). https://securingtomorrow.mcafee.com/consumer/consumer-threat-notices/most-hackable-gifts/

[10] Jim Donaldson. 2016. Why Retailers Are Turning To RFID For Loss Prevention. Website. (Aug. 2016). https://www.mojix.com/retailers-rfid-loss-prevention/

[11] The Daily Dose. 2017. Stopping Shoplifters Goes High-Tech. Website. (June 2017). http://www.ozy.com/fast-forward/stopping-shoplifters-goes-high-tech/78920

[12] Robert Epstein. 2016. The New Censorship. Website. (06 2016). https://www.usnews.com/opinion/articles/2016-06-22/google-is-the-worlds-biggest-censor-and-its-power-must-be-regulated

[13] National Retail Federation. 2017. The Economic Impact of the U.S. Retail Industry. Website. (2017). https://nrf.com/resources/retail-library/the-economic-impact-of-the-us-retail-industry

[14] FICO. 2017. Behavioral Analytics Attack Fraud, Cyber and Financial Crime. (04 2017). http://www.fico.com/en/blogs/analytics-optimization/behavioral-analytics-for-fraud-cyber-and-financial-crime/

[15] Frederic Filloux. 2013. Google News: the secret sauce. Website. (02 2013). https://www.theguardian.com/technology/2013/feb/25/1

[16] Arik Gabbai. 2015. Kevin Ashton Describes fithe Internet of Thingsfi. Magazine. (01 2015). https://www.smithsonianmag.com/innovation/kevin-ashton-describes-the-internet-of-things-180953749/

[17] Garmin. 2017. Garmin Edge 500. Website. (2017). https://buy.garmin.com/en-US/US/p/36728#overview

[18] Gartner. 2017. Gartner Says 8.4 Billion Connected "Things" Will Be in Use in 2017, Up 31 Percent From 2016. (02 2017).

[19] Christian Gilbertson. 2017. Here's How The Sensors in Autonomous Cars Work. Website. (03 2017). http://www.thedrive.com/tech/8657/heres-how-the-sensors-in-autonomous-cars-work

[20] Google. 2017. Mobile App Reporting in Google Analytics - iOS. Website. (2017). https://developers.google.com/analytics/devguides/collection/firebase/ios/#how_does_it_work

[21] Andrew Hawkins. 2017. Nvidia says its new supercomputer will enable the highest level of automated driving. Website. (10 2017). https://www.theverge.com/2017/10/10/16449416/nvidia-pegasus-self-driving-car-ai-robotaxi

[22] Patrick Hoge. 2010. Survey: 44% stop at Google News headlines. Website. (01 2010). https://www.bizjournals.com/sanfrancisco/stories/2010/01/18/daily24.html

[23] Nabeel Hyatt. 2017. Autonomous driving is here, and itfis going to change everything. Website. (04 2017). https://www.recode.net/2017/4/19/15364608/autonomous-self-driving-cars-impact-disruption-society-mobility

[24] IBM. 2015. IBM Plans to Acquire The Weather Companyfis Product and Technology Businesses; Extends Power of Watson to the Internet of Things. Press Release. (10 2015). http://www-03.ibm.com/press/us/en/pressrelease/47952.wss

[25] Ajit Jaokar. 2017. Behavioural Biometrics, IoT and AI. Website. (10 2017). https://www.datasciencecentral.com/profiles/blogs/behavioural-biometrics-iot-and-ai

[26] Search Engine Journal. 2016. Over 60% of People Trust Google for News vs. Actual News Sources. Website. (01 2016). https://www.searchenginejournal.com/google-news-2/154475/

[27] Jacob Kastrenakes. 2017. Most smart TVs are tracking you fi!? Vizio just got caught. (02 2017). https://www.theverge.com/2017/2/7/14527360/vizio-smart-tv-tracking-settlement-disable-settings

[28] Suzanne Kattau. 2015. Research from Gartner: Real-Time Analytics with the Internet of Things. Website. (06 2015). https://www.rtinsights.com/research-from-gartner-real-time-analytics-with-the-internet-of-things-dw/

[29] John Kiernan. 2017. Credit Card & Debit Card Fraud Statistics. Website. (02 2017). https://wallethub.com/edu/credit-debit-card-fraud-statistics/25725/

[30] Sam Levin and Mark Harris. 2017. The road ahead: self-driving cars on the brink of a revolution in California. Website. (03 2017). https://www.theguardian.com/technology/2017/mar/17/self-driving-cars-california-regulation-google-uber-tesla

[31] Karen Lewis. 2017. Visa and IBM are bringing the world secure payment experiences through the IoT. (02 2017). https://www.ibm.com/blogs/internet-of-things/visa/

[32] Annie Lin. 2017. Facial recognition is tracking customers as they shop in stores, tech company says. Website. (11 2017). https://www.cnbc.com/2017/11/23/facial-recognition-is-tracking-customers-as-they-shop-in-stores-tech-company-says.html

[33] Jon Markman. 2017. Amazon Using AI, Big Data To Accelerate Profits. Website. (06 2017). https://www.forbes.com/sites/jonmarkman/2017/06/05/amazon-using-ai-big-data-to-accelerate-profits/#12f2f9cb6d55
[34] Mark Mather. 2016. Fact Sheet: Aging in the United States. Media Guide. (01 2016). http://www.prb.org/Publications/Media-Guides/2016/aging-unitedstates-fact-sheet.aspx
[35] Kayla Matthews. 2017. 4 Statistics That Reveal Major Problems With IoT Security. Website. (02 2017). https://channels.theinnovationenterprise.com/articles/4-statistics-that-reveal-major-problems-with-iot-security
[36] Neil McAllister. 2014. How many mobile apps collect data on users? Oh ... nearly all of them. Website. (02 2014). https://www.theregister.co.uk/2014/02/21/appthority_app_privacy_study/
[37] Microsoft. 2017. Five ways edge computing will transform business. Website. (09 2017). https://blogs.microsoft.com/iot/2017/09/19/five-ways-edge-computing-will-transform-business/
[38] Patrick Nelson. 2016. Just one autonomous car will use 4,000 GB of data/day. Website. (12 2016). https://www.networkworld.com/article/3147892/internet/one-autonomous-car-will-use-4000-gb-of-dataday.html
[39] University of Missouri System. 2016. Securing the Internet of Things (IoT). Website. (11 2016). https://www.umsystem.edu/makeitsafe/securing_the_internet_of_things_iot
[40] Dan O'Shea. 2017. Report: Walmart developing facial-recognition tech. Website. (07 2017). https://www.retaildive.com/news/report-walmart-developing-facial-recognition-tech/447478/
[41] Kasey Panetta. 2017. Gartner Top Strategic Predictions for 2018 and Beyond. Website. (10 2017). https://www.gartner.com/smarterwithgartner/gartner-top-strategic-predictions-for-2018-and-beyond/
[42] Nevon Projects. 2017. IOT Heart Attack Detection & Heart Rate Monitor. Website. (2017). http://nevonprojects.com/iot-heart-attack-detection-heart-rate-monitor/
[43] Lopez Research. 2013. An Introduction to the Internet of Things (IoT). Research Report. (11 2013). https://www.cisco.com/c/dam/en_us/solutions/trends/iot/introduction_to_IoT_november.pdf
[44] Symantec Security Response. 2016. Mirai: what you need to know about the botnet behind recent major DDoS attacks. Website. (10 2016). https://www.symantec.com/connect/blogs/mirai-what-you-need-know-about-botnet-behind-recent-major-ddos-attacks
[45] Freddie Roberts. 2017. Half of US companies hit by IoT security breaches, says survey. (06 2017). https://internetofbusiness.com/half-us-iot-security-breach/
[46] Brenda Salinas. 2013. High-End Stores Use Facial Recognition Tools To Spot VIPs. Website. (07 2013).
[47] Antony Savvas. 2014. The Weather Company turns to open source big data analytics. Website. (11 2014). https://www.computerworlduk.com/data/kpmg-launches-big-data-investment-fund-3489089/
[48] Claire Scholz. 2015. Biometrics to Secure the Internet of Things. Website. (12 2015). https://blog.bioconnect.com/2552/biometrics-to-secure-the-internet-of-things/
[49] Stilgherrian. 2013. Smart TVs are dumb, and so are we. Website. (10 2013). http://www.zdnet.com/article/smart-tvs-are-dumb-and-so-are-we/
[50] Mark Sullivan. 2012. A brief history of GPS. Website. (08 2012). https://www.pcworld.com/article/2000276/a-brief-history-of-gps.html
[51] Raj Talluri. 2017. Why edge computing is critical for the IoT. Website. (10 2017). https://www.networkworld.com/article/3234708/internet-of-things/why-edge-computing-is-critical-for-the-iot.html
[52] Versa Technology. 2017. How much Data will The Internet of Things (IoT) Generate by 2020? Website. (10 2017). https://www.versatek.com/blog/how-much-data-will-the-internet-of-things-iot-generate-by-2020/
[53] Heather Thompson. 2017. Edge computing: Itfis what healthcare IoT craves. Website. (03 2017). http://www.medicaldesignandoutsourcing.com/edge-computing-healthcare-iot-craves/
[54] Weather Underground. 2017. Weather Underground - About Our Data. Website. (2017). https://www.wunderground.com/about/data
[55] Jack Vaughan. 2013. Google's big data infrastructure: Don't try this at home? Website. (10 2013). http://searchdatamanagement.techtarget.com/opinion/Googles-big-data-infrastructure-Dont-try-this-at-home
[56] Steven J. Vaughan-Nichols. 2017. fiQuHow to keep your smart TV from spying on you. Website. (03 2017). http://www.zdnet.com/article/how-to-keep-your-smart-tv-from-spying-on-you/
[57] Visa. 2017. Visa Ready and IoT Payments. Website. (2017). https://usa.visa.com/visa-everywhere/innovation/visa-ready-and-iot-payments.html
[58] Visa. 2017. Visa Ready: Biometrics. Website. (2017). https://visaready.visa.com/Biometric_program_detail.html
[59] Harro Ten Wolde and Eric Auchard. 2014. Germany's top publisher bows to Google in news licensing row. Website. (11 2014). https://www.reuters.com/article/us-google-axel-sprngr/germanys-top-publisher-bows-to-google-in-news-licensing-row-idUSKBN0IP1YT20141105
[60] Tim Worstall. 2014. If Google News Is Worth $100 Million Then Why Can't Google Pay The Newspaper Publishers? Website. (12 2014). https://www.forbes.com/sites/timworstall/2014/12/14/if-google-news-is-worth-100-million-then-why-cant-google-pay-the-newspaper-publishers/#7496b2b555a1

B BIBTEX ISSUES

C ISSUES

DONE:

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

C.1 Assignment Submission Issues

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

C.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

C.3 Formatting

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

Other formatting issues

C.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}`

C.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

C.6 Character 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.

If you see a ffigure and not a figure in text you copied from a text that has the fi combined as a single character

C.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 inffate your paper if you are below the page limit

C.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 ffoat. 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}
```

re