Snitch Application:

Addressing Cyber Trust in Future Living Spaces

8/6/20

Coastal Virginia Center for Cyber Innovation

Russell Moore (Intern)

Dr. Krzysztof Rechowicz (Mentor)

Dr. Saikou Diallo (Mentor)

The future of smart homes is filled with excitement and ample opportunity to live in an environment that features extraordinary convenience and energy efficiency. Devices such as Amazon Alexa and Roku Smart TV's feature voice interaction options that provide the user with enhanced functionality. While voice-activated devices are accommodating, it is imperative to acknowledge how they work on the backend. The problem is that these devices collect and share a large amount of data from users who are oftentimes unaware that it's even happening. The Snitch App is being developed in order to inform the user of what exactly is at risk in the event their smart home device is subject to a data breach. Considering how prevalent identity theft is in the cyber world, it is important to recognize the levels of trust people have with their connected in-home devices and inform them of the design and disclosure of the device in order to create a more inclusive and just experience for user-device interaction.

To give a brief overview of the Snitch App research project, it focused on the back-end functionality of the application. The initial step of the research consisted of reading through user agreement forms of various smart devices and identifying how information is worded as well as where it was physically located in the document. Using that knowledge, I created an algorithm that displays the trust features of a specific device to the user. After completing both processes, I was able to understand how much sensitive information is at risk when interacting with smart devices as well as how to extract that information and display it in an accessible manner.

Scanning the WIFI Network

The Snitch App is personalized for the user's home, based on their connected devices. Scanning the WIFI network for those devices is a crucial step because it provides the MAC Address. A MAC Address is a unique identifier that gives the physical address of a computer as well as the name of its vendor. For the Snitch App, each vendor's name is needed in order to determine their corresponding trust features, which can then be presented to the user.

Reading the Legal Documents

Manually reading through lengthy legal documents for numerous smart devices, such as Roku, Nest, Amazon, etc., is a portion of the research that I considered to be the "dirty work." Throughout this analysis stage, I was searching for types of information regarding data collection/sharing and information access/security. I discovered that most companies bury the trust features of their devices among less important information, forcing me to have to sift through the entirety of the document to find what I was looking for. For example, in Amazon Alexa's Terms of Use, I discovered that interactions with Alexa "… may be stored on servers outside the country which you live." There may be different laws governing the saved data that may have to be deciphered case by case, which can be compared to people legally being able to gamble online even in a state where gambling is illegal. Once I discovered what the trust features were for each device, I then had to identify the location trend of the information. Amazon's legal documents in particular typically included information about data usage and security in the top third of the page. It is important to take note of this because it streamlines my algorithm. Instead of having to search through the entire document for trust features, it designates a part of the text where the information I want is located.

Topic Modeling

While beginning the process of topic modeling, I knew I needed to have local access to each legal document. My initial approach consisted of a web scraping technique that I quickly discovered was not usable since companies such as Roku and Amazon, have security features built into their websites that prevent web scraping. After realizing this, I resorted to downloading the files into a local database, which brought me to the next step of cleaning the data. By tokenizing and lemmatizing the text followed by removing punctuation and stop words, I had prepared the data for topic modeling.

The previous step of manually reading through each form was necessary so I could understand the logic of the document, which I referred to when going through the topic modeling stage to discover hidden structures. For the snitch App, I determined that Latent Dirichlet Allocation (LDA) Topic Modeling in particular would be best because it provides an in-depth statistical model for certain topics appearing in each document. By assigning topics to an arrangement of words, LDA Topic Modeling helps map the document.

User Interface

Now that I had the information I was looking for; it is necessary to consider how it will be displayed to the user. Since the Snitch App must be accessible for people who are on the vision, physical, cognition, and autism spectrum, it is imperative to include features such as a colorblind-friendly color palette and a text to speech function. Complying with accessibility standards is a priority for the Snitch app, which will improve navigation as well as the user's overall understanding.

Throughout my research this summer I have greatly improved my knowledge of how smart-home devices operate and the importance of keeping sensitive and personal information secure. While there is work to be continued for the Snitch App, I have a firm understanding of the levels of trust people have with their connected in-home devices and a solid technical foundation to build off of. It is my goal to commercialize the Snitch App at the conclusion of its development.