**Malware Detection Presentation Script:**

Joe =   
George =   
Ruben =

Time: 30 minutes or less

Team introduction: names/ project name

**Joe**: Why did we choose malware detection?

* Big cyber security issue in our current time, machine learning is a great way to use new methods to combat old problems.

**Joe**: Background on the data?

* We got the data from a cyber security analysis… (a bit more on this topic)
* We also had to find a parser to suit the data from (source) which took most exe files and outputted a text of hex values that we are using as our parameters.

**Ruben**: Description of presentation

* Walk through the presentation: We designed a machine learning algorithm using python’s modules SCI-Kit/SK learn in jupyter notebook. We tested our algorithm with our databank with over 140k lines and over 60 features. And got good results for recall, precision and F-score. In this presentation we will walk you through our code structure and do a demonstration of a random EXE file at the end.

**George**: Walk through machine learning code

* Open jupyter notebook code file: Talk a little bit about the imports, SK learn
* We started by reading the file and separating it, the databank initially has a legitimate column which signifies if its malware or not, we used this to break it into two groups. We used random forest classifier to test and train the data. We also dropped many features to outlines which one has the most significant weights. (talk about other classifiers)
* Show the output of our initial data in whatever classifier

**Joe:** Talk about the parser

* Open up the parser in terminal and talk about the background and how to build it.
* Show a sample output of the parser like njRAT.txt

**Ruben:** Talk about parser of the parser

* In order to use this data we need to first parse the output of the parse. This Is where we get our first issue, the output gives us about 30-33 features instead of the full 60 in the original databank. Fortunately, the first 34 feature matched identically to our databank. By using python, we wrote a quick function to take the output of the parser and give a viable output we can use in our machine learning, which we will show you in a bit.

**George:** Demonstration

* Here is a short demonstration of our algorithm. Pick a ramdon.exe and run it through the parser, parser of parser.py and run the machine learning algorithm. Talk about the process and explain what the output means.