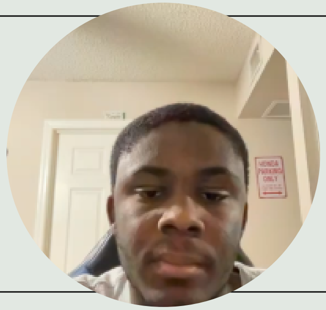
Big Data in Malware Detection

Queenly Xie, Russell Ridley, Lakshmi Katravulapalli

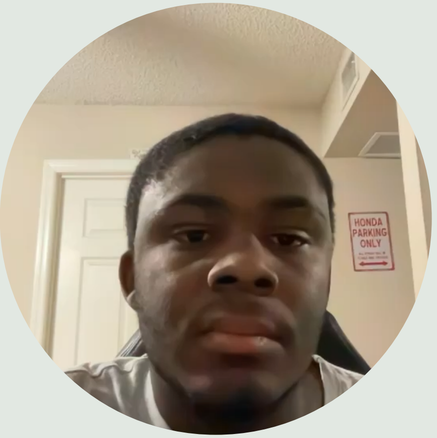
Department of Electrical and Computer Engineering, University of Central Florida

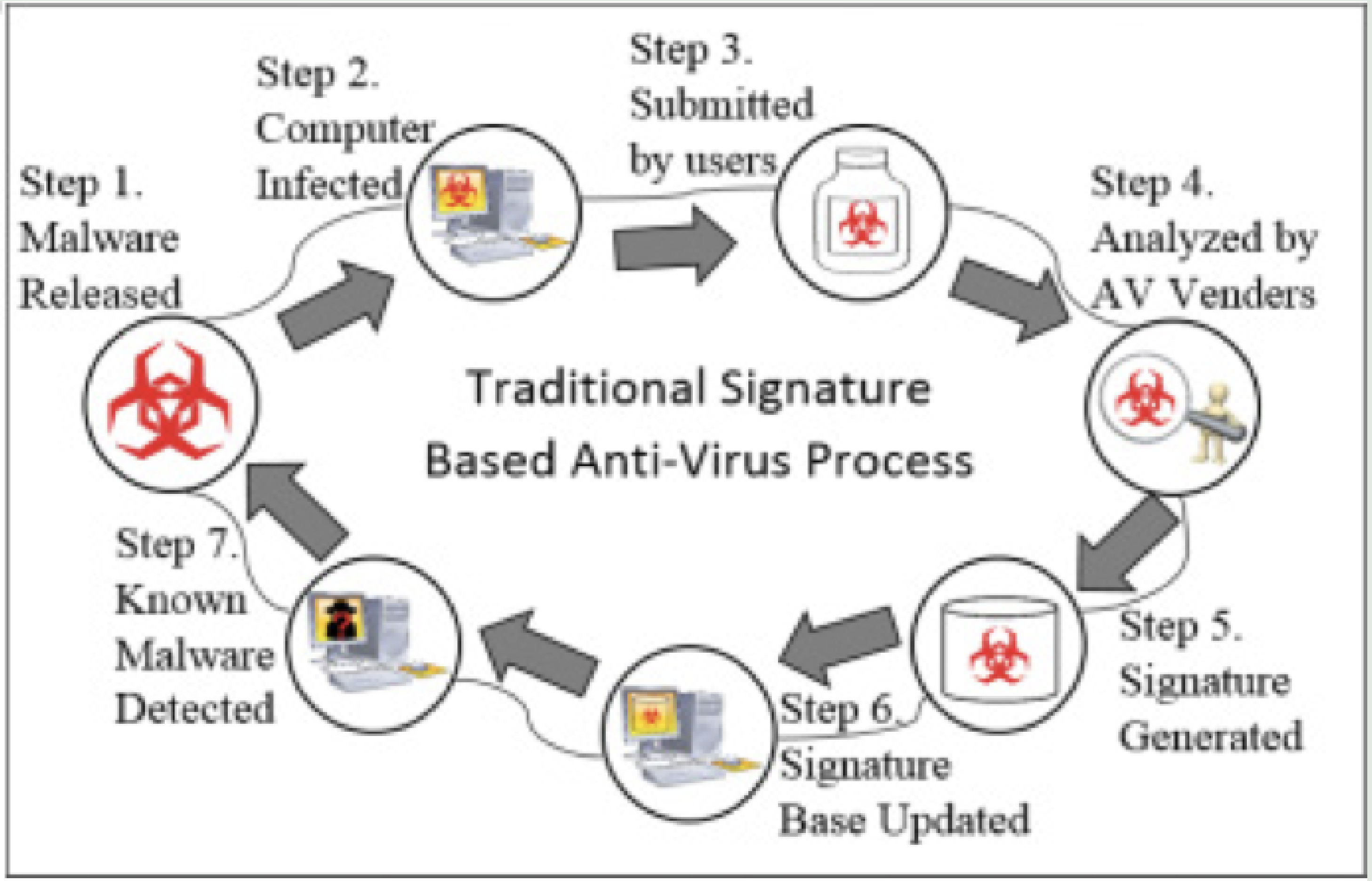
Background



- **Prevalence**: Malware impacts millions of devices, performing actions like data leaks, file encryption, and personal losses.
- **Importance**: Critical for protecting computers and mobile devices.
- **Objective**: Explore and outline various strategies for enhancing malware detection and prevention, drawing insights from existing research.

Tradition al Malware Detection




Step 1. Malware Released
Step 2. Computer Infected
Step 3. Submitted by users
Step 4. Analyzed by AV Venders
Step 5. Signature Generated
Step 6. Signature Base Updated
Step 7. Known Malware Detected
Traditional Signature Based Anti-Virus Process

Y. Ye, T. Li, D. Adjeroh, and S. S. Iyengar, “A Survey on Malware Detection Using Data Mining Techniques,” ACM Computing Surveys, vol. 50, no. 3, pp. 1–40, Jun. 2017, doi: https://doi.org/10.1145/3073559.

Algorithms/Models

Community Detection:

- Community detection allows us to discover hidden patterns and structures within a specified system or network. This helps speed up the detection process and identify malware that might have similar attack methods.

Data Mining:

- Data Mining has the ability to analyze larger sets of data and use intelligent methods such as Feature Extraction and Clustering to help detect signs of Malware.

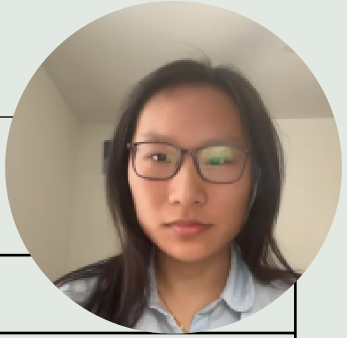
Deep Learning Algorithms:

- Deep Learning Algorithms present the ability to create scalable and sophisticated malware detection models that can handle large datasets and adapt to changing environments.

Android Malware Detection with DBN:

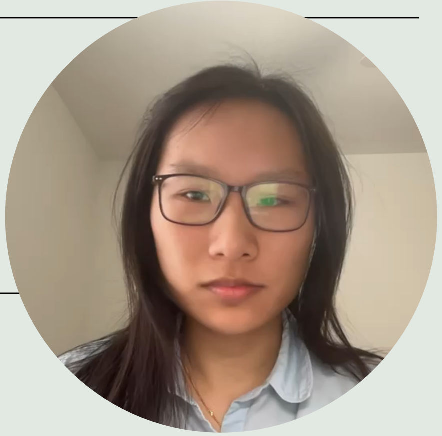
- DBN is an artificial neural network used for Android Malware detection. This system can classify android applications based on recognized patterns to determine if they are safe or harmful.

Deep Learning & Across Industry



	About	**Advantage**	**Drawback**
Deep Transfer Learning	➢ Pre-trained deep learning models employed on non-malware situations for adaption to malware detection.	- Applications on cloud server/platform - Lightweight models on resource-constraint mobile devices - Pre-train models - Low computational/memory overhead	- Domains for pre-trained model and executed domain must be similar - Recognition of new malware attacks - Black box interpretability
Artificial Neural Networks (ANNs):	➢ ANN is used to examine features that are extracted from files and data, determine if they are benign or malicious based on pattern recognition.	- Feature Learning - Scalability - Real-Time detection - Adversarial Robustness - Deep Learning Architectures	- Data dependency - Privacy Concerns - Extra Resources - Overfitting
Generative Adversarial Networks:	➢ Uses generative algorithms to find new patterns from inputted data.	- Enhancing training data - Feature Extraction - Adversarial Training - Anomaly Detection - Generate simulated malware threats	- Training instability - Overfitting - Challenges when evaluating bigger sets of data
Mobile Edge Computing	➢ Improves mobile networks' efficiency and specifically for mobile devices, improve security and defense.	- Intrusion Detection Systems - Intrusion Prevention Systems - Real-time detection	- Reliance on network - Integration challenges
Adversarial Machine Learning for Intrusive Detection Systems (IDS)	➢ Machine learning based security mechanisms that addresses malware attacks that evade detection.	- Online intrusion detection systems - Proactive defense - Adaptability	- Computational/resource overhead - Needs to account for false malware inputs

Themes:



Online Platforms: Web-based services

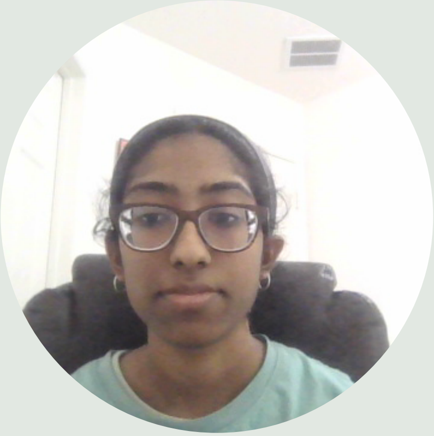
Mobile Computing: Mobile devices that offer portability

Inter-field Applications:

- Hardware based malware detection applied to mobile computing.
- Machine learning based detection for online platforms and operating systems (MAC OS, autonomous vehicles).

Conclusion

We have reviewed and presented the different type of Malware detection algorithms and techniques. One of the challenges presented is the increase in Adversarial attacks which are known to trick classifiers by changing data distribution. Researchers are developing new techniques to fight misleading instances and make them stronger against these attacks. Researchers plan to focus more on Android and Desktop Platforms hoping to create specialized models to target these online platforms.



Contribution

Russell Ridley:

- Completed slides 2 and 3 on the lightning presentation.
- Researched the current problems with malware detection algorithms.
- Researched ways companies some malware detection algorithms to protect customer data.

Lakshmi Katravulapalli:

- Researched and organized different algorithms and models used for Malware detection gathered from research papers.
- Completed slide 4 and 8 on the lightning presentation
- Worked on the conclusion and identified some of the challenges and future objectives of Malware detection

Queenly Xie:

- Read and compiled papers on malware detection specifically on online platforms.
- Completed slide 5 and 6 on the lightning presentation.
- Create graphics and figures grouping and comparing malware detection and prevention techniques.
- Contribute ideas of possible areas of improvement and future areas of research/methods to analyze more accurate the malware detection.