Collaborative Research: SaTC: CORE: Medium: Defending Against Social Engineering Attacks with In-Browser Al

- Roberto Perdisci Lead PI (<u>perdisci@uga.edu</u>) Nick Nikiforakis – PI (<u>nick@cs.stonybrook.edu</u>)
- Phani Vadrevu PI (phani@cs.uno.edu)

NSF Grants No. 2126641, 2126654, 2126655









Challenge

- Social engineering (SE) attacks exploit humans' decision-making process.
- They use a large variety of deception and persuasion tactics. They are at the basis of many web-based scams, phishing attacks, and malware distribution. They cause significant monetary loss. • Existing defenses (e.g., URL blockers) are mostly ineffective.

SE attack examples





Technical Solution + Impact

- SE Scout: automatically discover and
- collect datasets of attack examples.
- **SE Learn**: Learn how to detect future SE

	virus have been detected	Amazon Rewards Event	
oŋ	Wour iDhono		
Value	iOS System Alert	Congratulations Amazon user!	
beca visit pho the p	Your iPhone is infected with virus and immediate urgent action is required.	We would like to thank you for your loyalty to Amazon - so we are offering you the chance to receive a reward. Spin the wheel to claim your special prize []	
4 m		Good Luck!	
How Step	Continue and follow the instructions to fix your iPhone.	ОК	
App virus	**Exit at your own risk**	(2)	
Step	OK	CO TO	
	Remove Virus Now		

attacks using multiple ML models.

- SE Shield: Deploy efficient ML models within the browser.
- **Privacy**: ML models will run locally, without sharing users' browsing data.
- *Impact*: Block SE attacks encountered by users while browsing.

SE Sco	SE Learn			SE Shield			
Deep Reinforcement Learning Agents		SE Attacks	Deep	<section-header></section-header>	Models	Model Compression	In-Browser Al Defenses
	Learning	Dataset	victore i formatione i formatio	<text><text><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></text></text>	Multi-Modal Detection	Efficient Detection	





Fully-Convolutional Network



Broader Impact

• Educating senior citizens: Senior citizens are one of the main targets of SE attacks. We plan to work with AARP and other organizations to educate senior citizens on how to identify and defend themselves against web-based scams.

The 5th NSF Secure and Trustworthy Cyberspace Principal Investigator Meeting

June 1-2, 2022 | Arlington, Virginia