AMERICAN UNIVERSITY OF BEIRUT

AI for Cyber Security: The Good and the Bad

Imad H. Elhajj
American University of Beirut

Workshop on Building Trust in Digital Government Services

11 September 2023

Assumptions

TRUST

DIGITAL

GOVERNMENT

SERVICES

What is Trust?

believe in the **reliability, truth, or ability** of …

Security is a state of well-being of information and infrastructures in which the **possibility** of **successful** yet **undetected** theft, tampering, and disruption of information and services is **kept low or tolerable**

Concerns?

Telecommunication

Banking and Finance

Government Services

Transportation

Medical Devices

Stuxnet

Highly specialized malware discovered July 2010

Solely targeting:

- SCADA systems
- Siemens SIMATIC WinCC
- SIMATICSTEP 7 software for process visualization and system control

Exploits a total of four unpatched Microsoft vulnerabilities (Two that **had yet to be disclosed**)

Compromises two digital certificates

Fingerprints industrial control systems to limit impact!

http://www.symantec.com/security_response/writeup.jsp?docid=2010-071400-3123-99

Gauss

Complex cyber-espionage toolkit Discovered June 2012

Functions:

- Intercept browser cookies and passwords.
- Harvest and send system configuration data to attackers.
- Infect USB sticks with a data stealing module.
- List the content of the system drives and folders.
- Steal credentials for various banking systems in the **Middle East**.
- Hijack account information for social network, email and IM accounts.

http://www.securelist.com/en/downloads/vlpdfs/kaspersky-lab-gauss.pdf

Gauss

Targets banking credentials

Vast majority of victims located in **Lebanon**

Gauss command-and-control (C&C) infrastructure was shutdown in July 2012

Nation-state sponsored attack?

Hacks of Cars

Ten car models (8 manufacturers)

Access all 10 and drove them away “by intercepting and relaying signals from the cars to their wireless keys”.

“The attack works no matter what cryptography and protocols the key and car use to communicate with each other.“

Equipment cost between $50 and $1000

Hacking Cars with Keyless Systems Feasible and Practical, Swiss Researchers Say, January 19, 2011, IEEE Spectrum

In January 2022, a 19-year-old researcher David Colombo revealed that he could exploit a bug in the TeslaMate dashboard to control over 25 vehicles in 13 different countries

Challenges

Communication Convergence

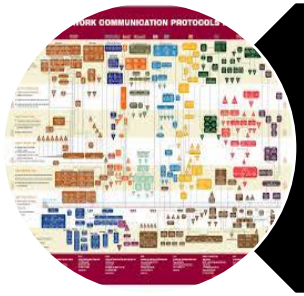
Scale

Specialized Connected Devices

Complex Software

Jurisdiction

Fundamental Vulnerabilities

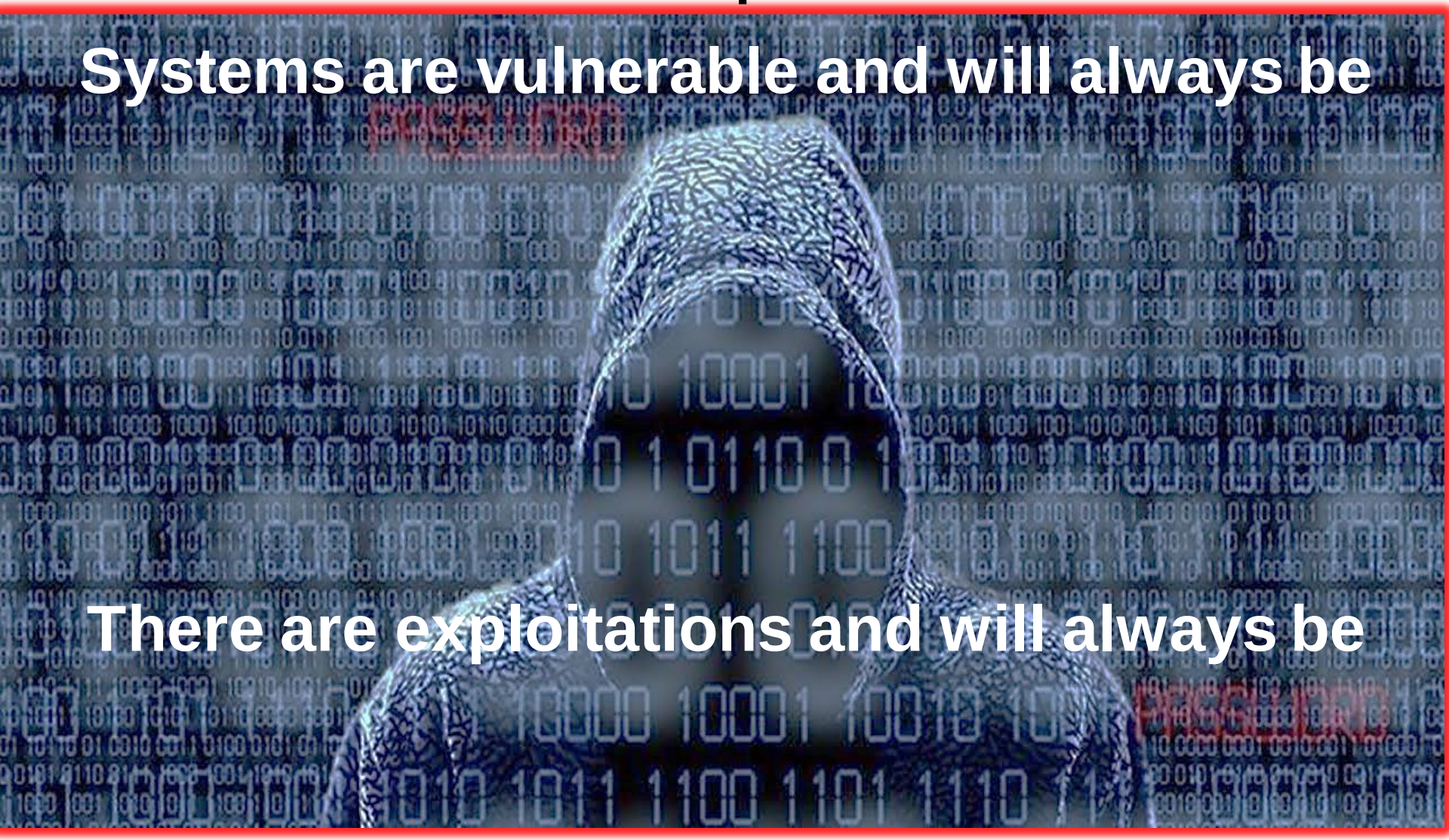


Protocols

Implementation

Services/Features

Assumptions



AI the Bad

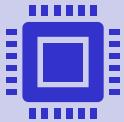
Generative AI

Spam
Avatars
Voice cloning

Automation of vulnerability scanning

Traffic spoofing

Examples



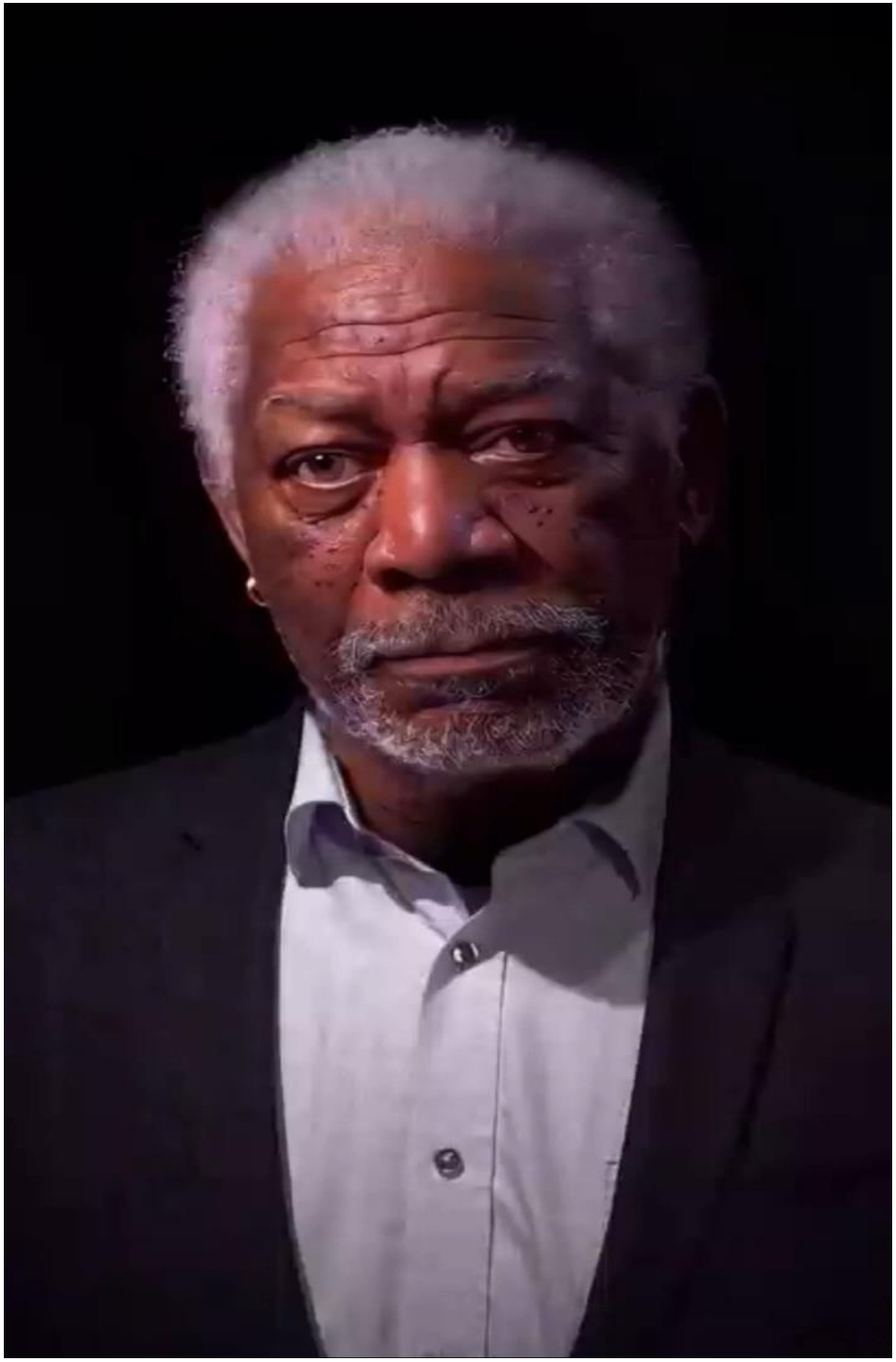
Security researchers forced Microsoft's Bing chatbot to behave like a scammer (https://greshake.github.io/

Deepfake (short segments of video needed)

Basic Deepfake (10,000 Iterations) - $20

High-Quality Deepfake (50,000 Iterations) - $80

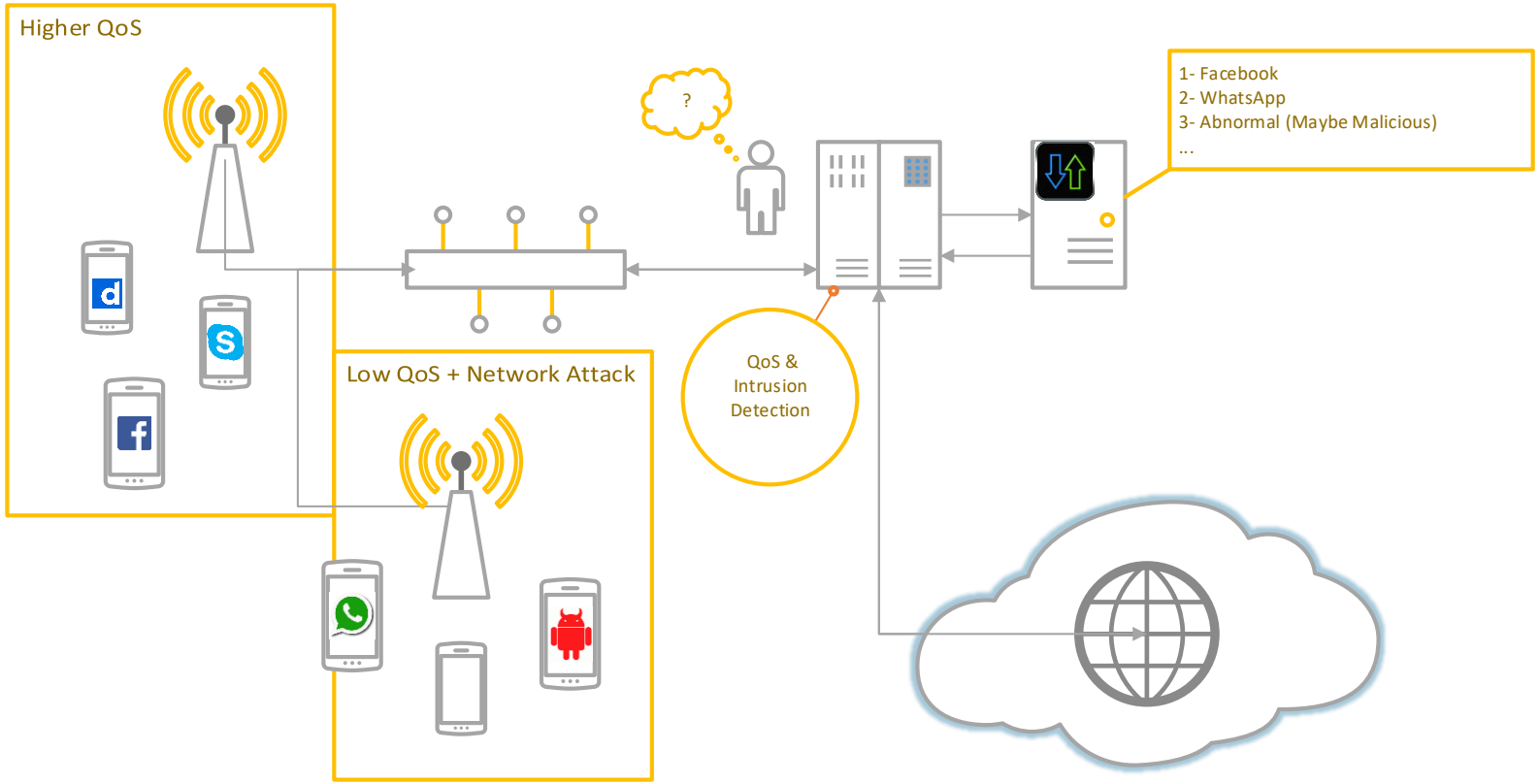
Voice cloning (short segments of voice needed)



AI the Good

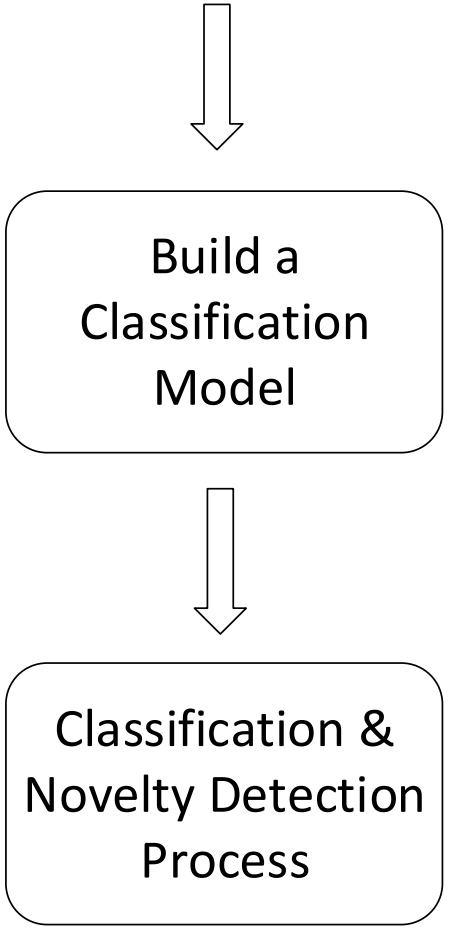
- Intrusion detection and prevention
- Spoofing detection
- Security orchestration
- Privacy enhancement
- Generative AI detection?

Traffic Classification and Intrusion Detection



Approach

- Does Not Require Deep Packet Inspection (DPI)
- Does Not Require installing clients (agents)
- Network side and non-invasive (transparent to intruders)
- Use of Machine Learning Algorithms



Results

- 96% Classification Accuracy
- 97% Detection accuracy for unknown traffic generated by benign Apps (Facebook, WhatsApp, …)
- 92% Detection accuracy for unknown traffic generated by malicious Apps (nMap, Packet Generator, …)
- Low False Alarm rate @ 3%
- Generalizability results not promising!

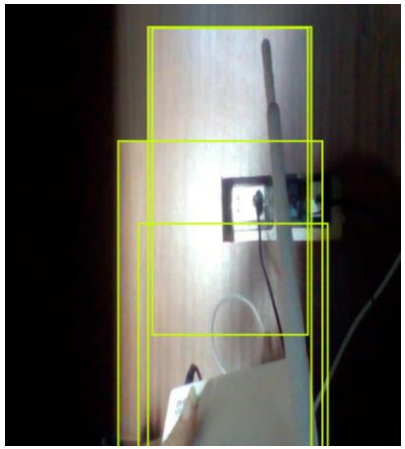
Real-time Detection of Assets

Motivation: Remote maintenance and support opens the door for privacy concerns

Balance between privacy and utility is needed

Client-side real-time object detection is needed

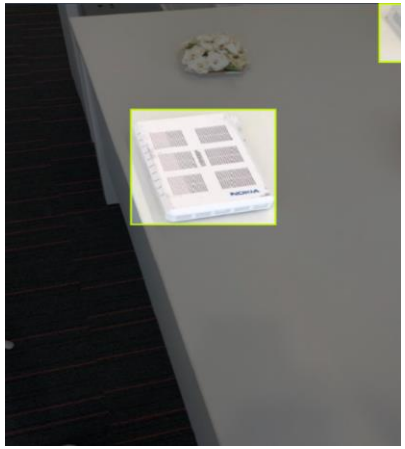
Develop dataset of CPEs and apply transfer learning



Yolov8

Yolov5

Yolov7-tiny



Yolov7

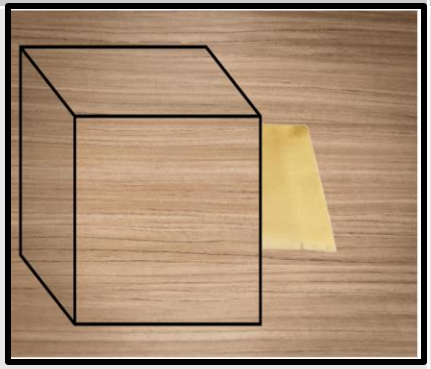
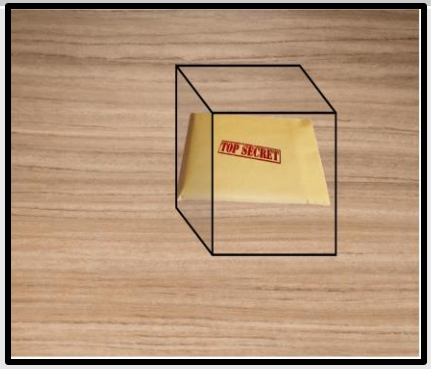
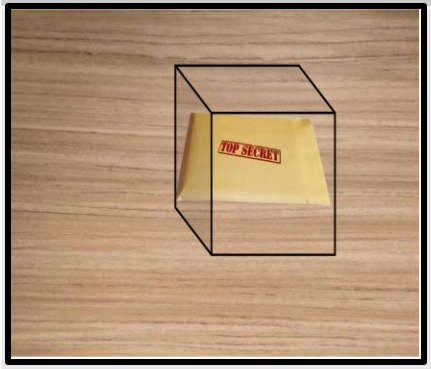
Preliminary Results

Model Version	Precision	Average Precision	Recall	Real-time Perfomance
Yolov5	90%	85%	75%	Inaccurate bounding boxes
Yolov7	**92%**	**87%**	**90%**	**Best Real-time performance (demo)**
Yolov7-tiny	98%	95.6%	92.3%	Bounding boxes rarely displayed
Yolov8	85.5%	92.9%	95%	Inaccurate bounding boxes

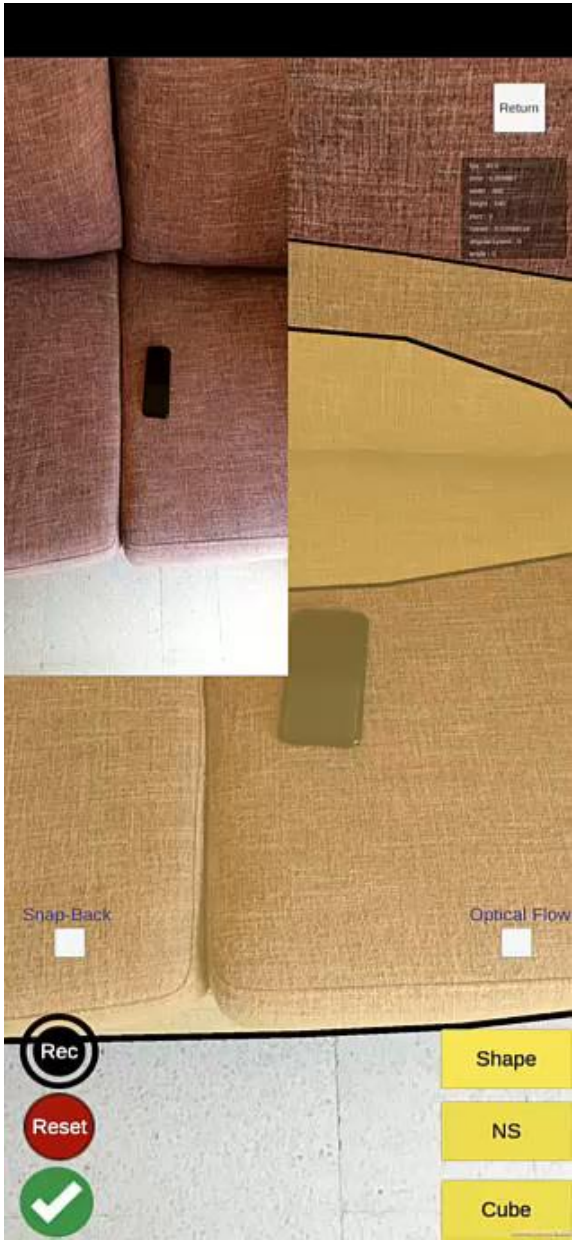
Router Detection Using YOLOv7 -no backend

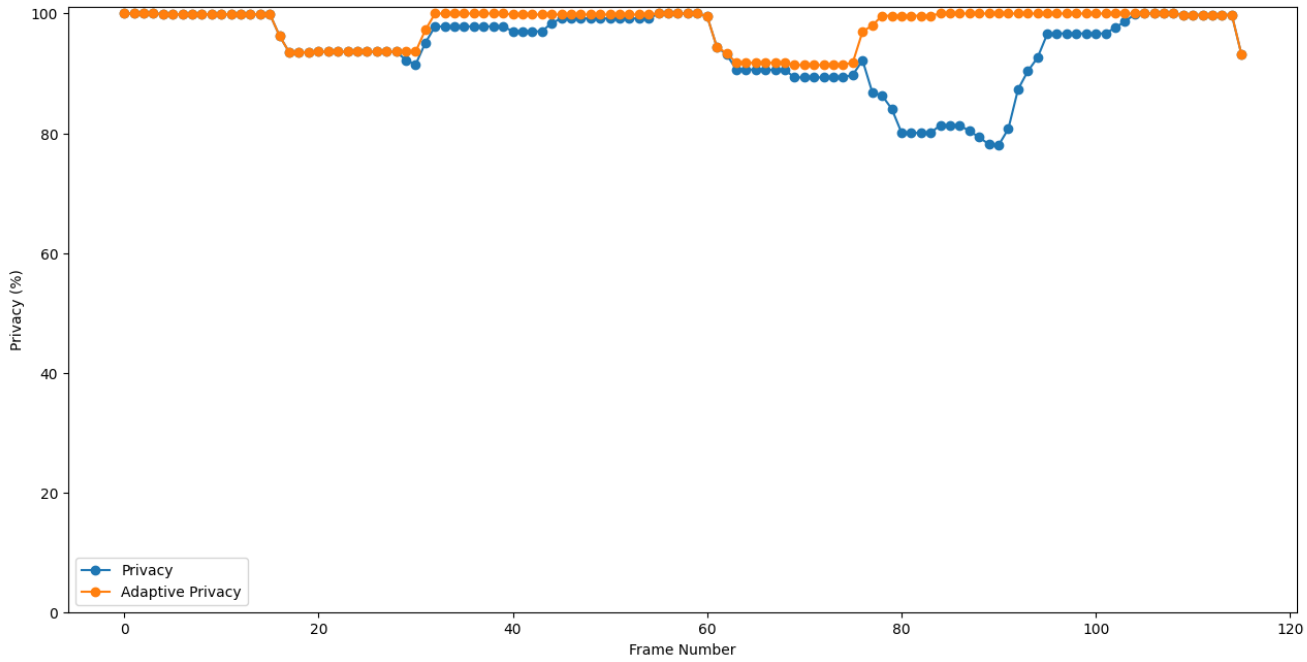
Mobile Diminished Reality for Preserving 3D Visual Privacy

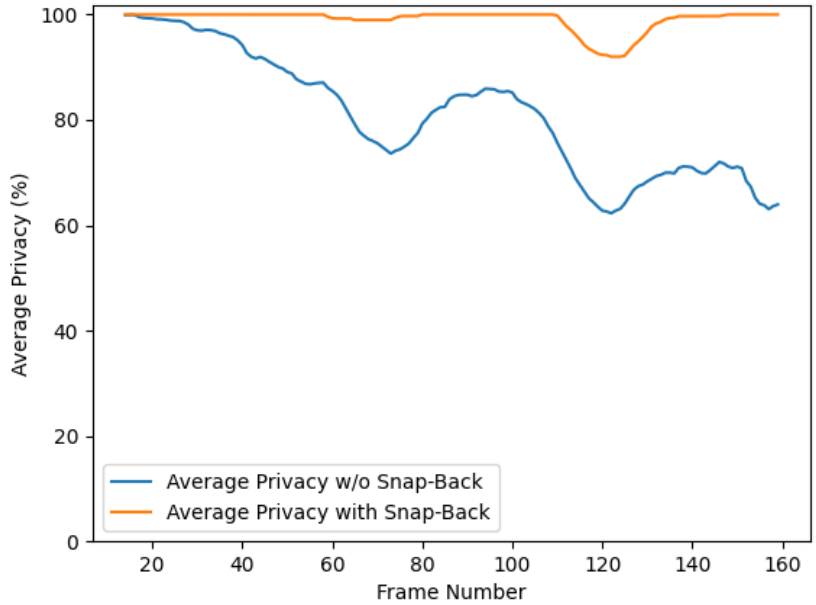
- **Motivation**: Privacy implications of Mixed Reality Applications
- **Solution:** We implement a *real-time privacy-preserving* Diminished Reality framework to obfuscate objects in *3D* while preserving the realism of the 3D environment

Features			
Obfuscation by Inpainting	Snap-Back	Optical Flow	Adaptive Privacy
			

Demo




Privacy (%)
Frame Number
Privacy
Adaptive Privacy
0
20
40
60
80
100
120


Average Privacy (%)
Frame Number
Average Privacy w/o Snap-Back
Average Privacy with Snap-Back
0
20
40
60
80
100
120
140
160

Final Thoughts

Exciting times

Great potential for innovation

Caution needed

Data governance

Identity management

AI governance

Thank you

ie05@aub.edu.lb

Acknowledgments:

- TELUS Corp. Canada
- CISCO
- AUB University Research Board