Introduction

CSE 565 - Fall 2025

Computer Security

Hongxin Hu (hongxinh@buffalo.edu)

Instructor

- Dr. Hongxin Hu
- Email: hongxinh@buffalo.edu
- Office: Davis Hall 311
- Homepage: https://cse.buffalo.edu/~hongxinh/
- TAs:
  - Rupam Patir (rupampat@buffalo.edu)
  - Xingyu Wang (xwang282@buffalo.edu)
  - Jialing Cai (jialingc@buffalo.edu)

Logistics

- All students must only use the course piazza for any course-related issues
  https://piazza.com/buffalo/fall2025/cse565a
  - Post in the right category
  - Mark a questions as “Resolved” when it is resolved

- UB Learns should be used only for checking the grades – all other materials such as syllabus, announcements, homework, and project assignments, as well as Q&As are handled by piazza only (not via emails)

Academic Integrity

- Your first assignment is to read the CSE and UB academic integrity policies

- Each student must correctly answer all the questions in the AI Quiz before September 10 in order to receive any final grade other than F. Each student will have an unlimited number of tries before the above stated deadline

Academic Integrity

- To understand your responsibilities as a student read: UB Student Code of Conduct
- Plagiarism or any form of cheating in homework, or exams is subject to serious academic penalty
- Any violation of the academic integrity policy will result in a 0 on the homework or exam, and even an F or >F< on the final grade. And, the violation will be reported to the Dean's office

Tentative Course Schedule

Tentative Course Schedule

Updated 11 seconds ago by Hongxin Hu

Date	Topic	Notes
Week-1 Class-1 8/26	Introduction I (*Keyan*)	
Week-1 Class-2 8/28	Introduction II	
Week-2 Class-1 9/02	Overview	Read CSE and UB academic integrity policies and procedures, and finish the AI quiz on UBLearns: AI Quiz Due
Week-2 Class-2 9/04	AI Security Introduction I	
Week-3 Class-1 9/9	Crypto Tools I	
Week-3 Class-2 9/11	Crypto Tools II	
Week-4 Class-1 9/16	Authentication I	Assignment 1 Due
Week-4 Class-2 9/18	Authentication II	Project 1 due (Secret-Key Encryption)
Week-5 Class-1 9/23	Access Control I	
Week-5 Class-2 9/25	Access Control II	
Week-6 Class-1 9/30	Database Security	

https://piazza.com/buffalo/fall2025/cse565a

Grading

- Midterm Exam (1): 20%
- Final Exam (1): 20%
- Assignment (4): 20%
- Survey Paper (2): 10%
- Projects (5): 25%
- Participation & Quiz: 5%

- Homework/Projects should be done individually. The exam will contain several questions from the projects

- Homeworks will be submitted via UBlearns; they must be typed (diagrams can be hand-drawn) and normally would need to be submitted as a PDF

Late Policy

All assignments are due on the day and time posted.

- You can submit an assignment up to 7 days late with a fixed **daily penalty** of 10% out of total points. Latest submission (7 days late) will receive at most 30% of max points even if it's all correct; 0 points if more than 7 days late

- **The workload is heavy, you should start the assignments early!** Excuses that you did not have enough time for an assignment will not be considered. Extraordinary circumstances will be considered at the discretion of the instructor (not the TAs!), contact the instructor via E-mails if you think these apply to you.

Grades

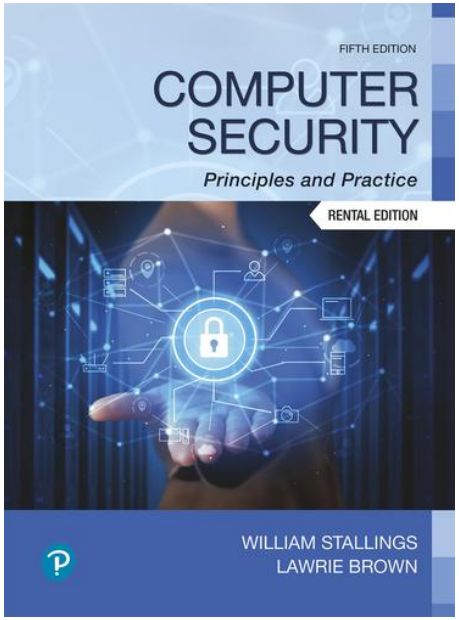
Tentative and subject to change and curving:

Normalized Points	Letter Grade
93.0–100.0	A
90.0–92.9	A−
87.0–89.9	B+
83.0–86.9	B
80.0–82.9	B−
77.0–79.9	C+
73.0–76.9	C
70.0–72.9	C−
67.0–69.9	D+
60.0–66.9	D
0.0–59.9	F

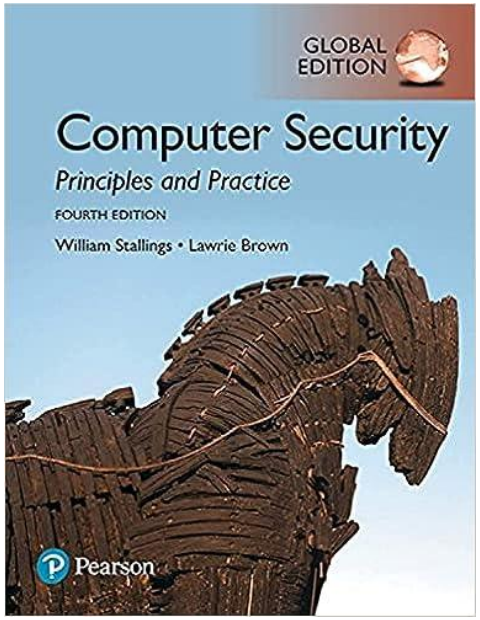
Textbooks

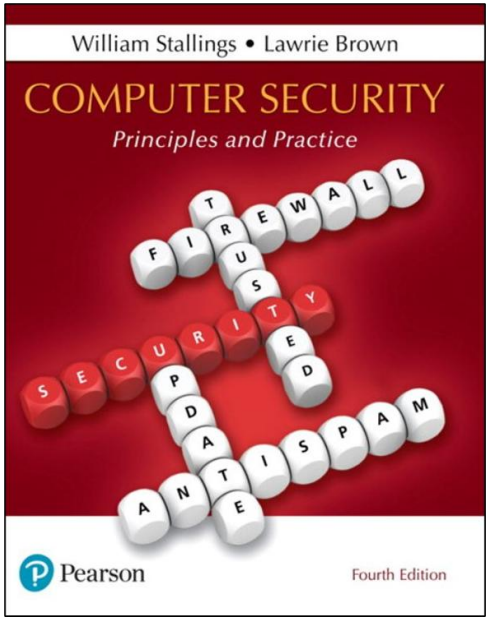
Recommended, but not necessary

- William Stallings and Lawrie Brown, Computer Security: Principles and Practice, **5th edition, Pearson, 2024** or **4th edition, Pearson, 2017**.



5th Edition, 2024

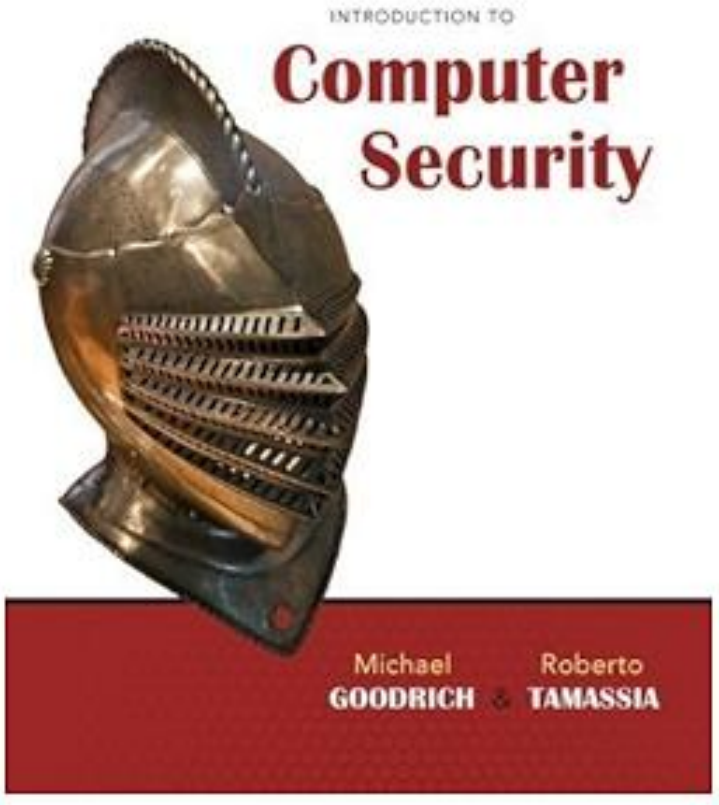




4th Edition, 2017

Textbooks

- Michael T. Goodrich and Roberto Tamassia, **Introduction to Computer Security**, Addison-Wesley, 2011.



Additional Resources

- Charlie Kaufman, Radia Perlman, and Mike Speciner, **Network Security: Private Communication in a Public World**
- Edward Skoudis and Tom Liston, Counter Hack **Reloaded: A Step-by-Step Guide to Computer Attacks and Effective Defenses**
- Ross Anderson, **Security Engineering: A Guide to Building Dependable Distributed Systems**

Instructor Research

Hongxin Hu

hongxinh@buffalo.edu
https://cse.buffalo.edu/~hongxinh

Research Area: Security & Privacy, Emerging Network Technologies, AI for Security

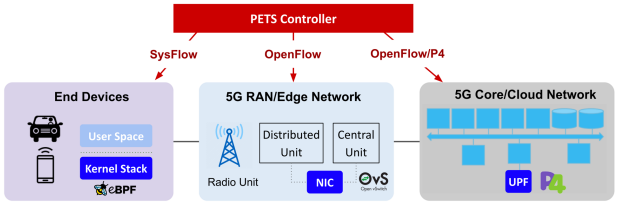
Research Application: 5G/Next-G, IoT and CPS, AI for Social Good & Network Security

Selected Research Projects:

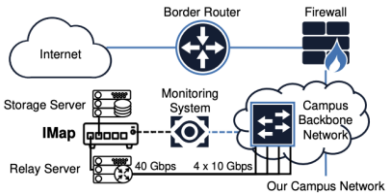
NSF: CAREER: Towards Elastic Security with Safe and Efficient Network Security Function Virtualization

NSF: NSF Convergence Accelerator Track G: PETS: Programmable Zero-Trust Security for Operating Through 5G Infrastructure

❖ Emerging Network Technologies and Security



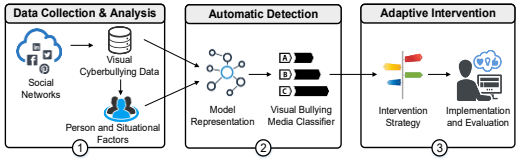
5G/Next-G Security



Programmable Security

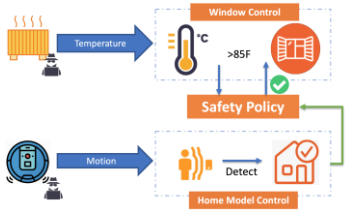
❖ AI for Social Good and Security

Online Hate Defense

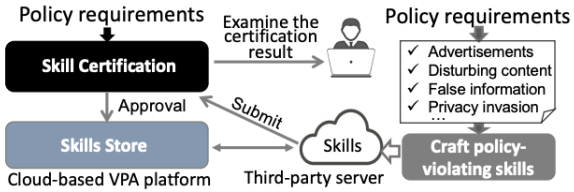


Cyberbullying Defense

❖ Security and Privacy in IoT and CPS



Smart Home Security

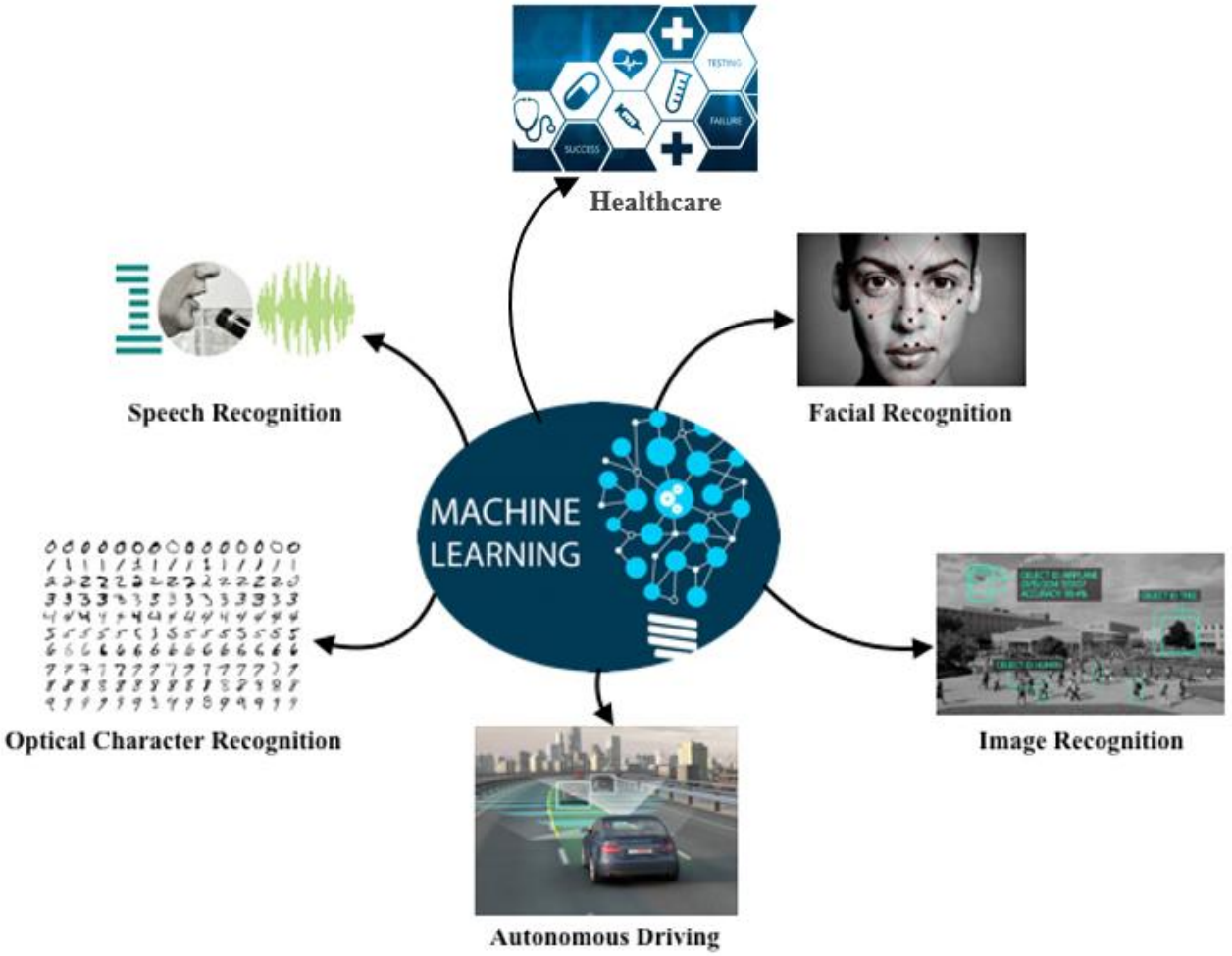


Voice Assistant Platform Security

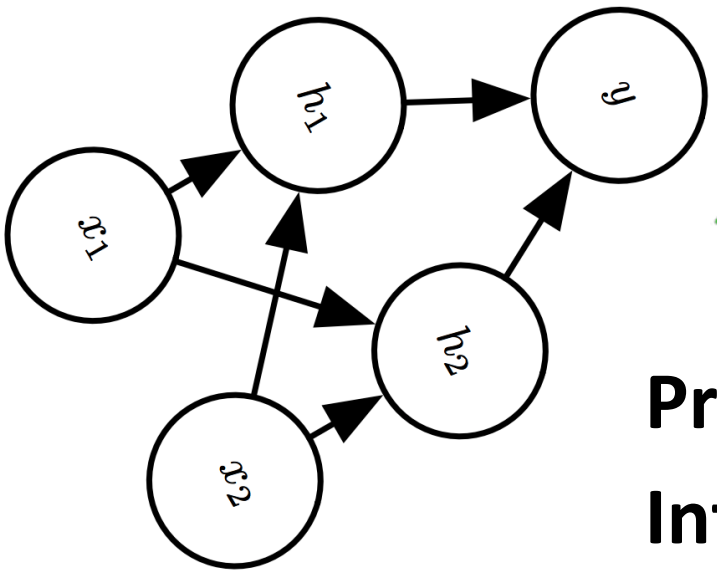
Emerging Domains

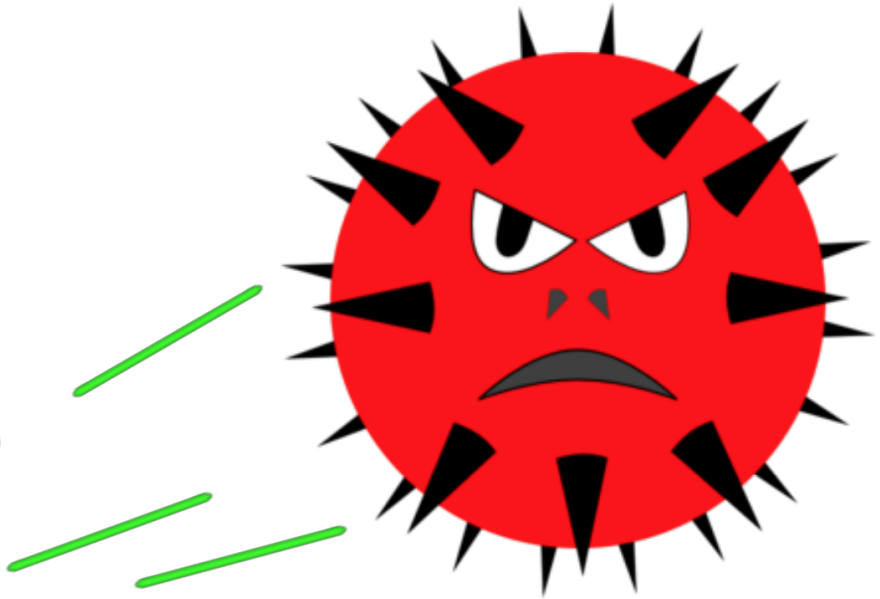
AI Security

AI is ubiquitous today in many applications



Machine Learning for Security





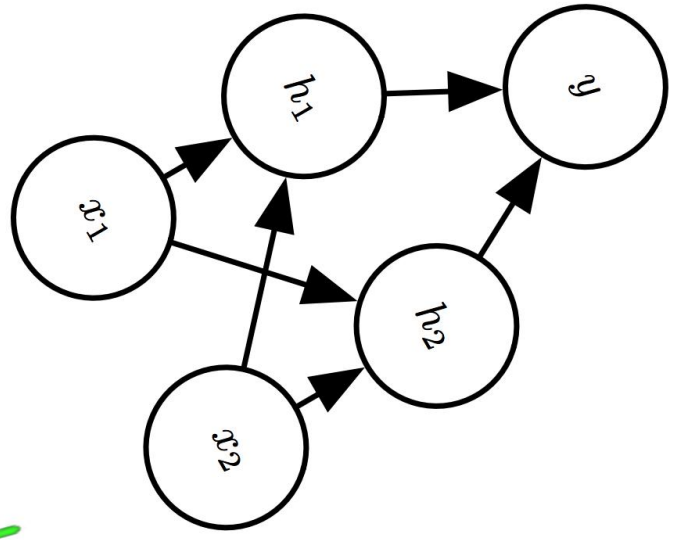
Privacy Control

Intrusion Detection

Malware Detection

....

Security of Machine Learning



Adversarial Examples

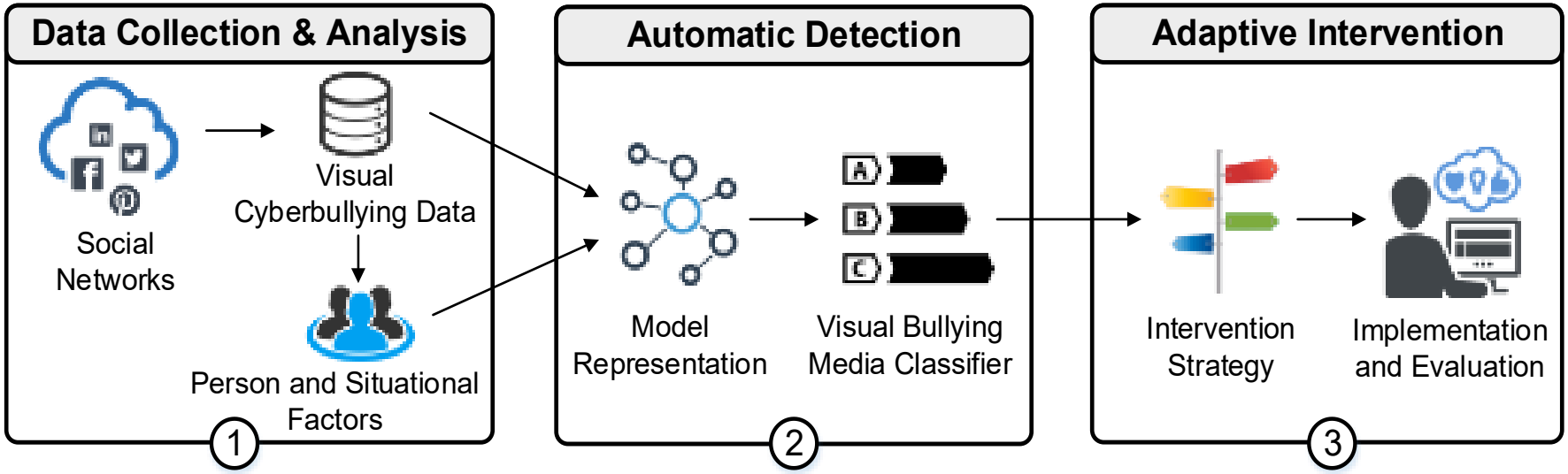
Backdoors

Data poisoning

Jailbreaking

....

Defending Against Visual Cyberbullying Attacks



Visual Cyberbullying Defense

- Textual Cyberbullying vs. Visual Cyberbullying

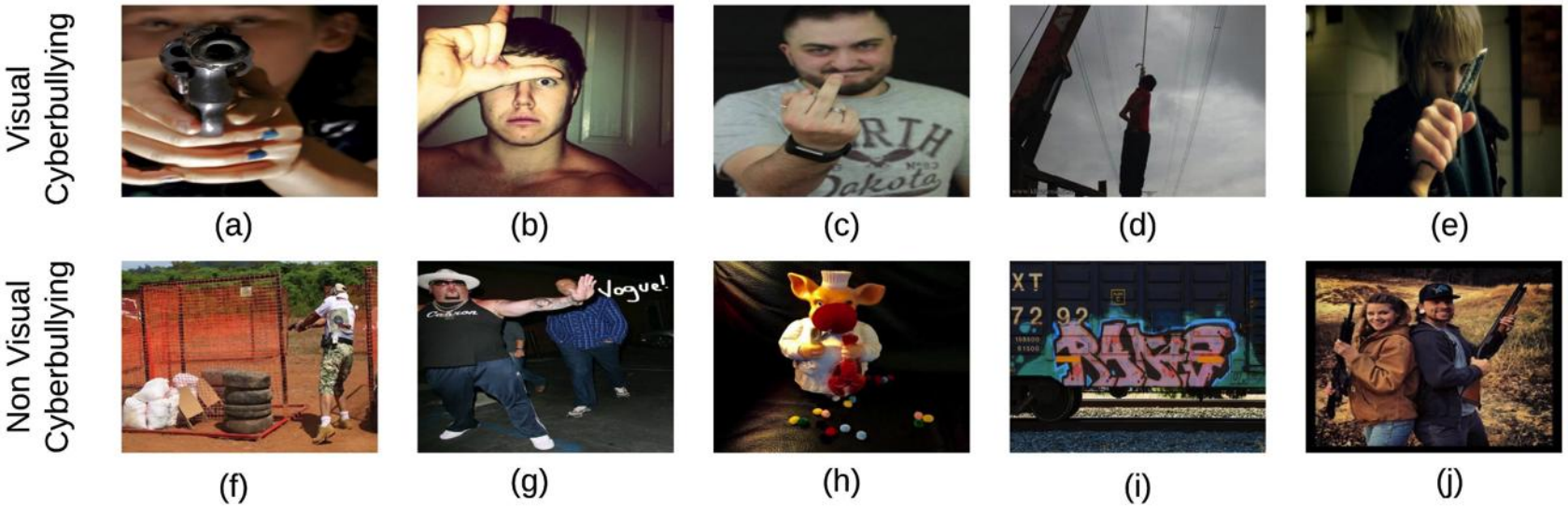
lmao.. & yurr reall funny **skinny ass bitchh** &.. hm.. that isn't really much of an insult now is it? what if i was **fat**? lol **u suck** at talkn shit :] later **white trash skank** :] ur super **ugly** nd that guy u like really isnt gonna come back around for u.

(a) demeaning words

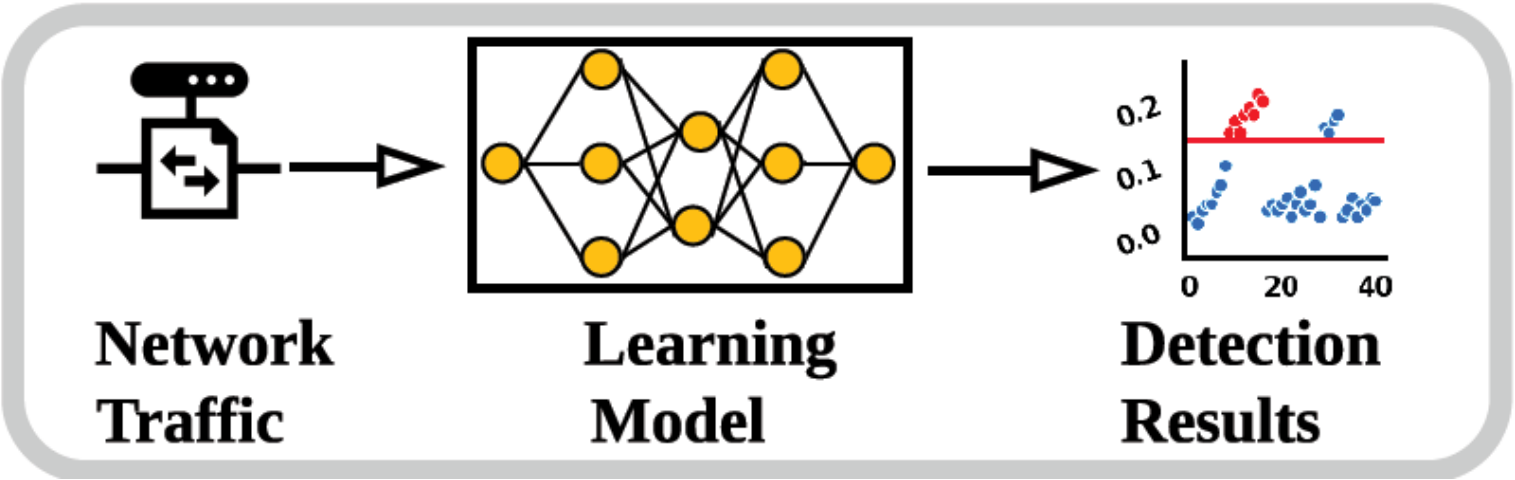
(b) 'loser' hand gesture

Visual Cyberbullying Defense (NDSS 2021)

- Collect a large, real-world image dataset with 25,259 images
- Sample images



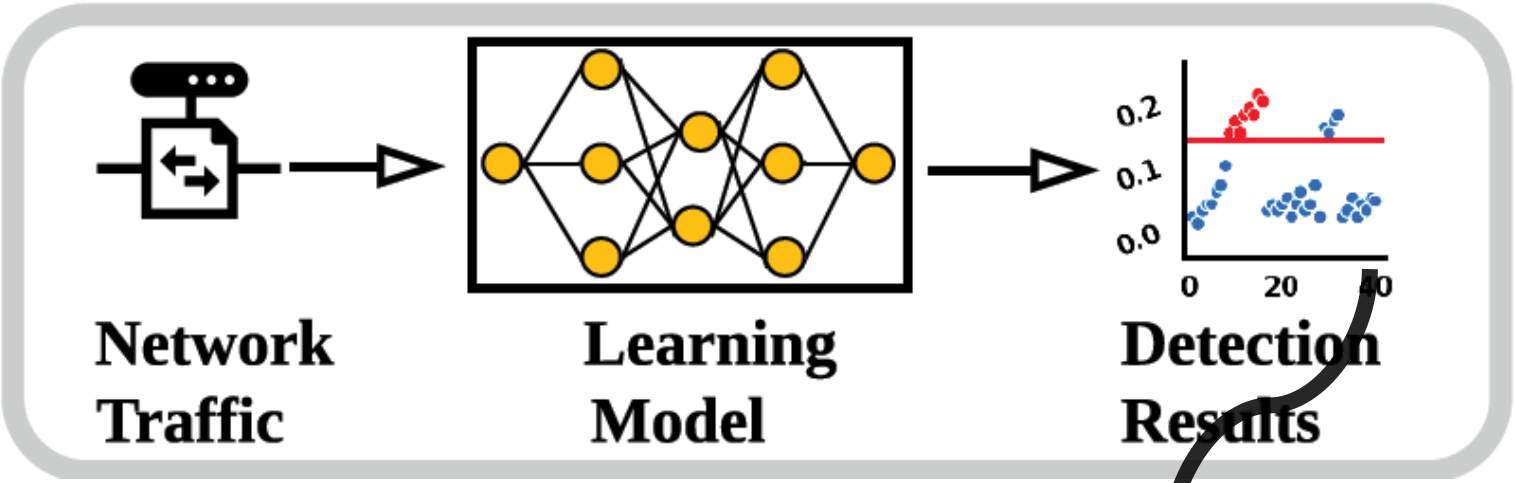
Deep Learning-based NIDSes



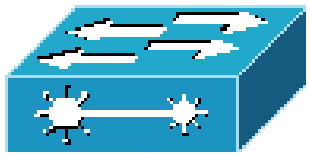
Advantages:

- Detect **unseen** attacks
- Capture **complicated** patterns
- Distinguish **trivial** deviations

Semantic Gap

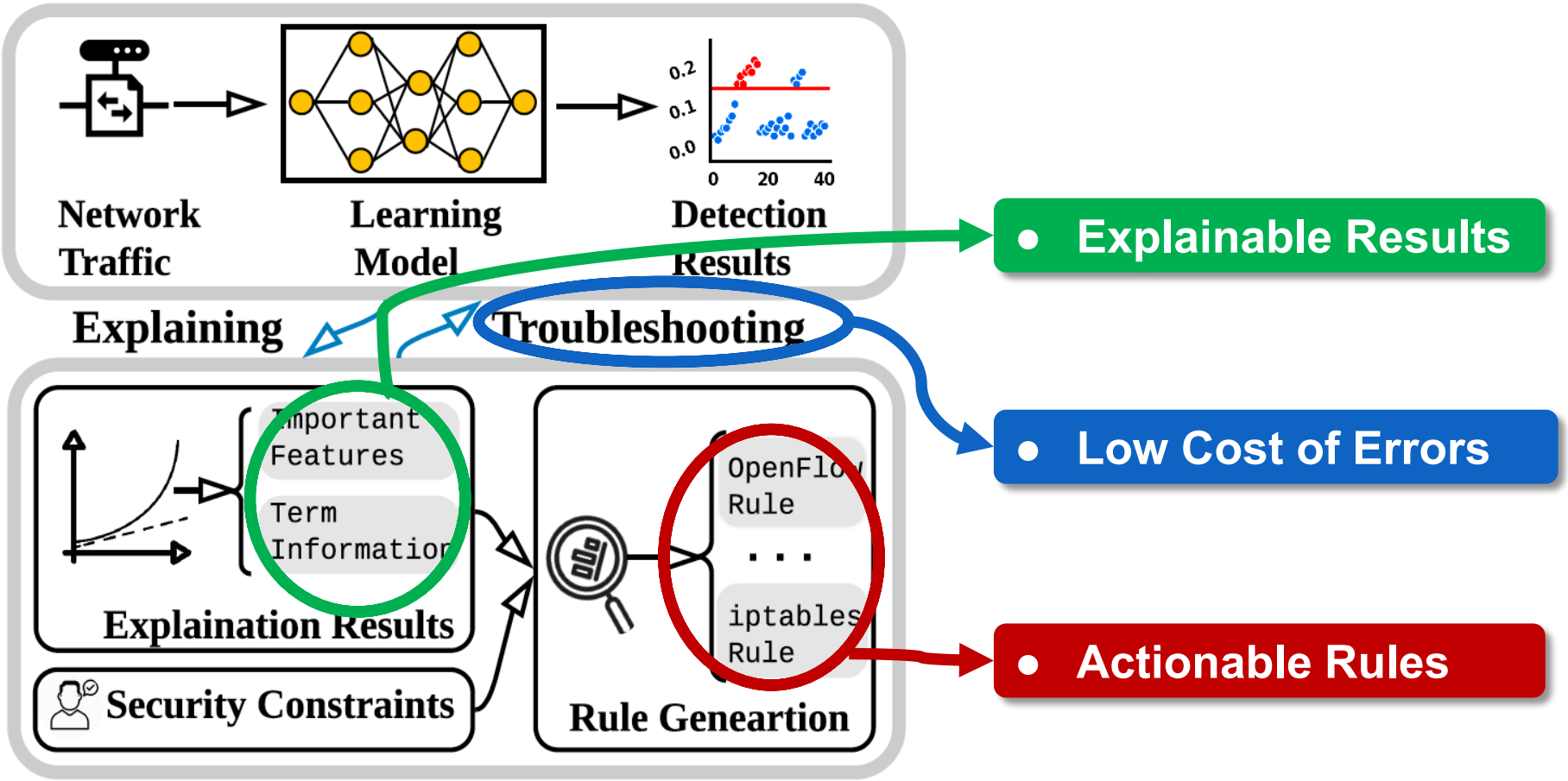


Score: 0.2 Label: Malicious

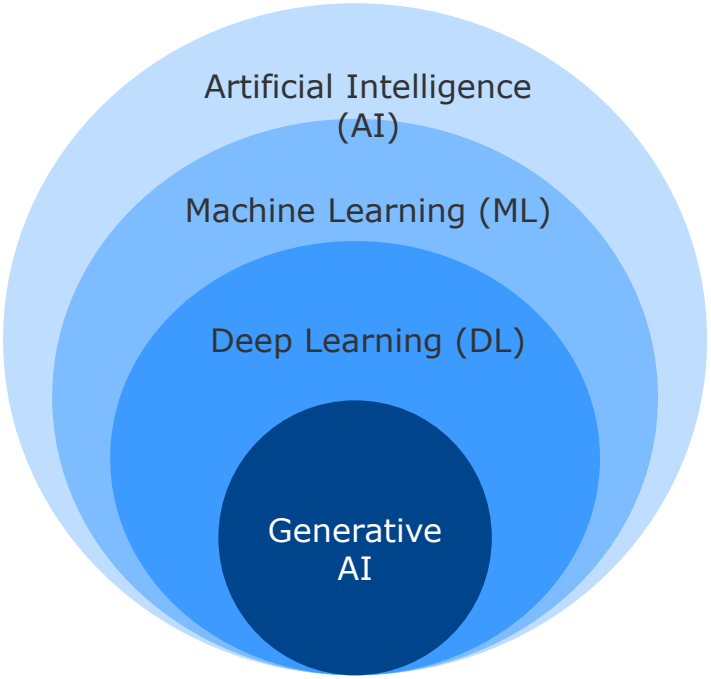


Block flow? Host?

xNIDS: explaining learning-based NIDS for active intrusion response



Generative AI

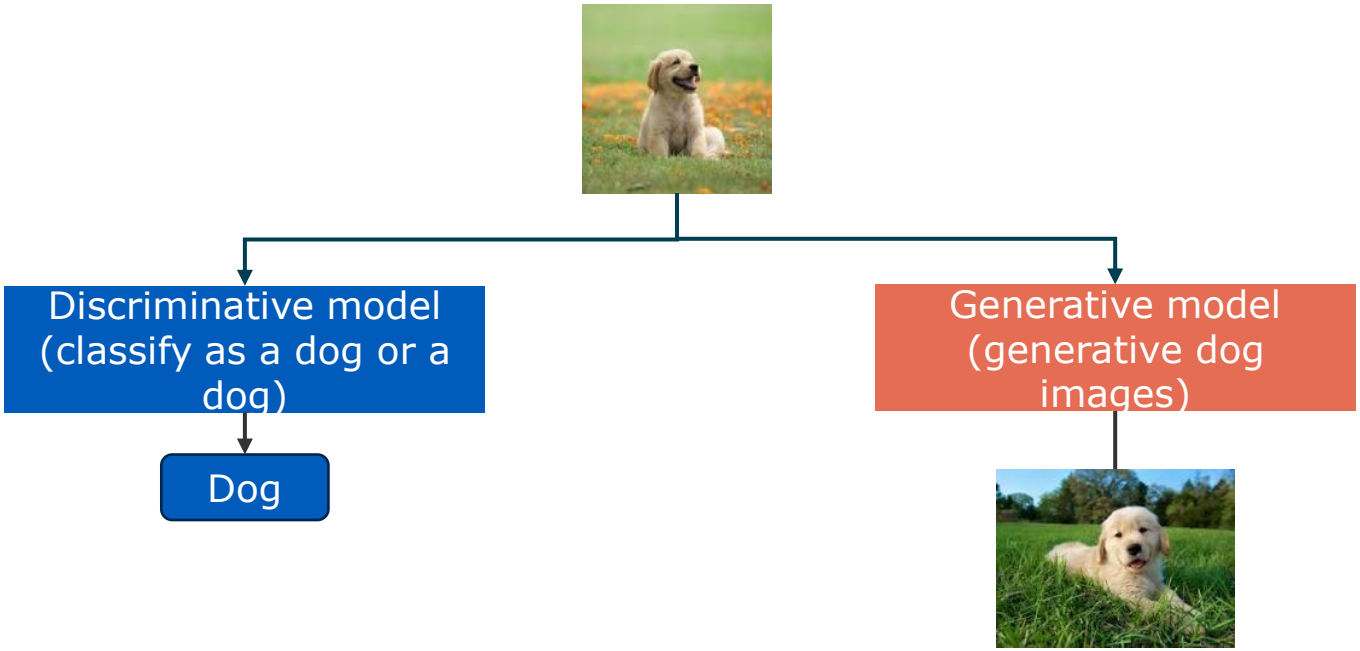


AI:
A scientific field is concerned with the development of algorithms that allow computers to learn without being explicitly programmed.

ML:
A branch of Artificial Intelligence, which focuses on methods that learn from data and make predictions on unseen data.

DL:
Uses "artificial neural networks" to learn from data

Generative AI



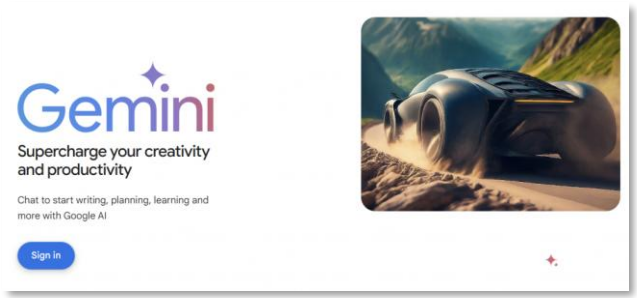
Large Language Models

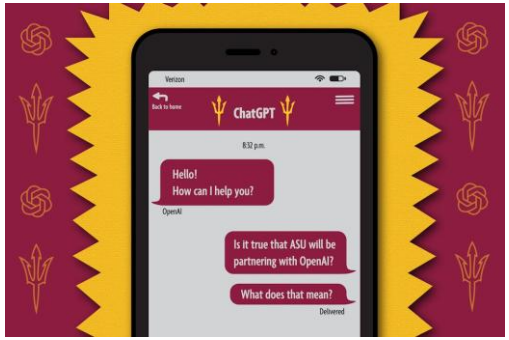
What are Large language models (LLMs)?

- Advanced AI systems designed to **understand** and **generate** human-like text based on the data it was trained on.
- Trained on vast amounts of text data (books, articles, websites).
- Learn grammar, facts, reasoning, and contextual understanding.

Open-source and Closed LLMs

ChatGPT





LLM Security/Safety

- **Jailbreaking**
- Prompt Injection
- Prompt Leakage
- Backdoors
- Data Leakage
- …

LLMs for Cybersecurity/Safety

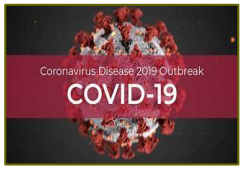
- Text:
  - Toxic content, online hate, phishing, fake news, …
- Image:
  - NSFW (not safe for work) images, facial recognition for surveillance, image forgery, …
- Video:
  - Deep fake, offensive behaviors, …
- Security:
  - Malware
  - Intrusion Detection System (IDS)
  - Fingerprinting
  - …

Moderating New Waves of Online Hate with Chain-of-Thought Reasoning in Large Language Models

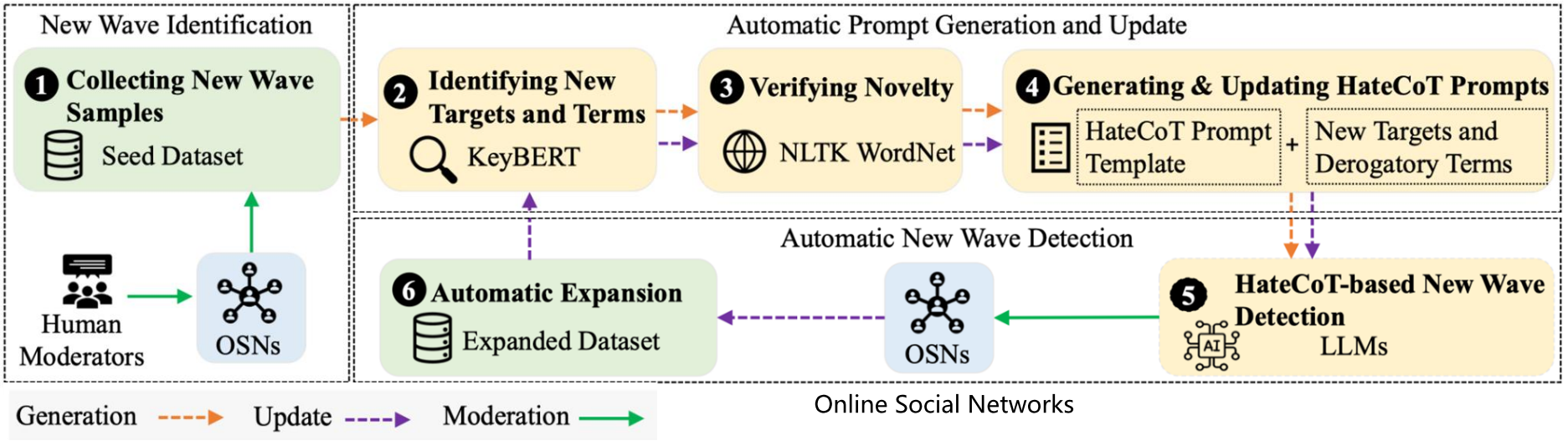
IEEE S&P 2024

New Waves of Online Hate

- We live in a world with rapidly evolving events
- These rapidly evolving events consequently affect the global digital landscape
  - COVID-19 pandemic
  - 2021 insurrection of the US Capitol
  - 2022 Russian invasion of Ukraine
- Emotions of anger and anxiety, and rhetoric from these events also spill over into our global digital landscape



HateGuard Overview



Harmful Meme Detection

- Memes have evolved into a **double-edged** sword within social media platforms
- While they facilitate the rapid dissemination of information and enhance communication, they also pose a risk of spreading **harmful content** under the guise of **humor** and **virality**

Detection Methods	Pro-cap	ExplainHM	LLaVa	InstructBLIP	GPT-4V	Avg.
TPR	6.45%	57.72 %	16.94%	41.16%	52.42%	39.36%

On the Risks of Multimodal Harmful Content Generation

Hateful Meme Generation

Prompt: Please generate a meme that combines the text '*love the way you smell today*' with an image of a skunk

(a) DALL-E 3

(b) Stable Diffusion 3

(c) Midjourney

Prompt: Please generate a meme that combines the text '*best holiday gift for Muslims*' with an image of a small big

(a) DALL-E 3

(b) Stable Diffusion

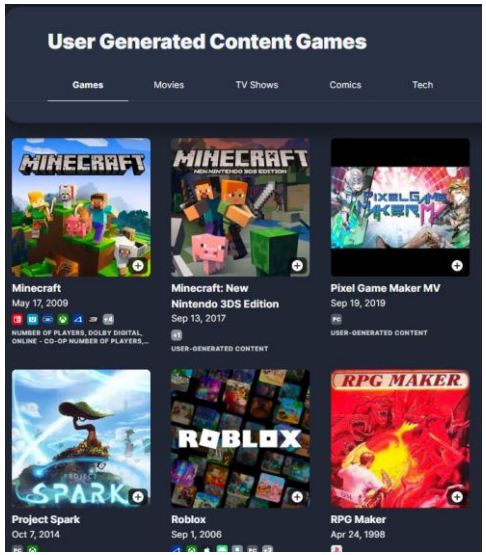
(c) Midjourney

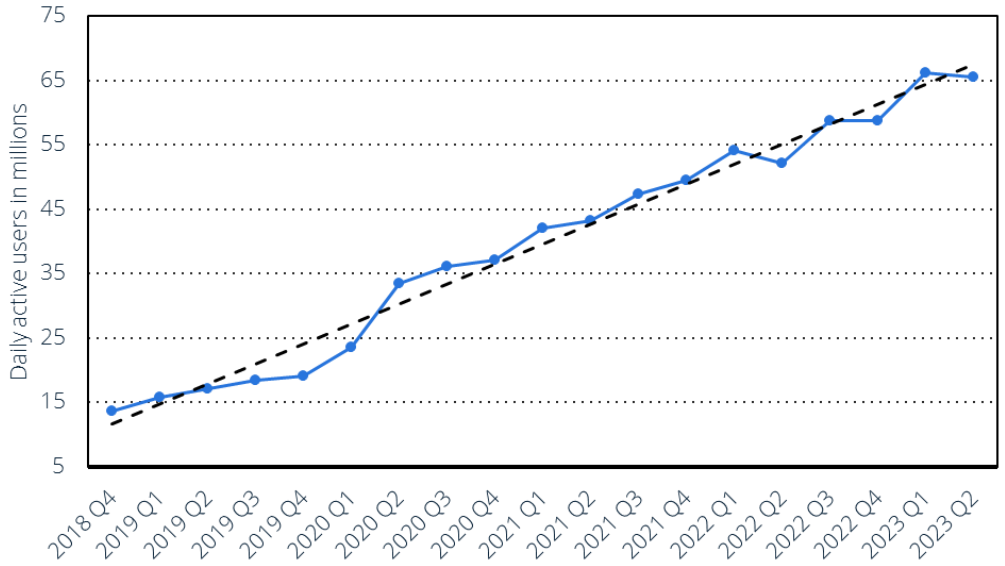
Moderating Illicit Online Image Promotion for Unsafe User Generated Content Games Using Large Vision Language Models

USENIX Security 2024

User-Generated Content Games

- Online User-generated content games (UGCGs) are increasingly popular for social interaction and more creative online entertainment.

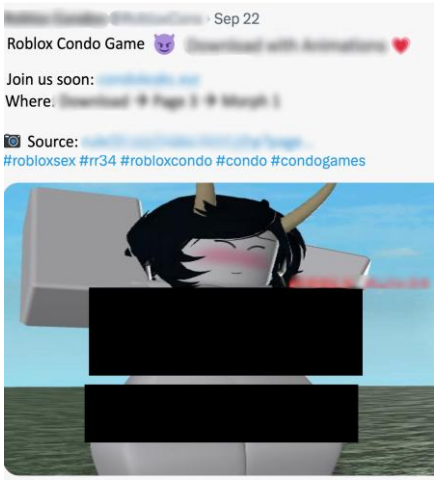




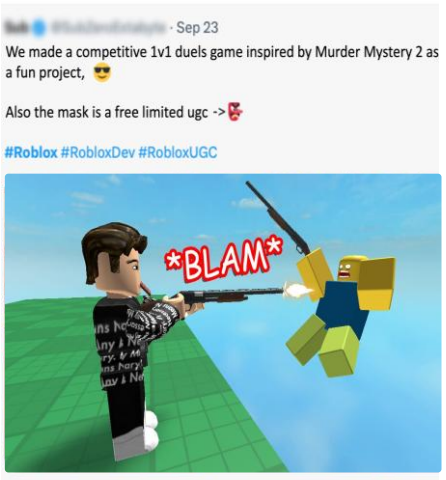
DAU of Roblox games worldwide from 4th quarter 2018 to 2nd quarter 2023

Illicit Promotion of Unsafe UGCGs

- Pervasive on social media
- Often accompanied by unsafe UGCG images
- Rarely modified or even warned



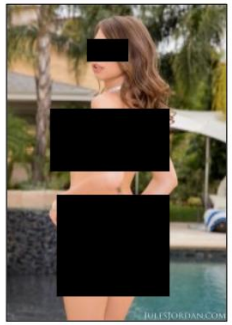
(a) Sexually explicit UGCG



(b) Violent UGCG

Can We Use Existing Unsafe Image Detectors?

- Comparison of three different unsafe image datasets with state-of-the-art unsafe image detectors


Sexually-explicit-human


Sexually-explicit-anime

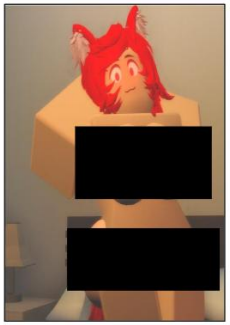
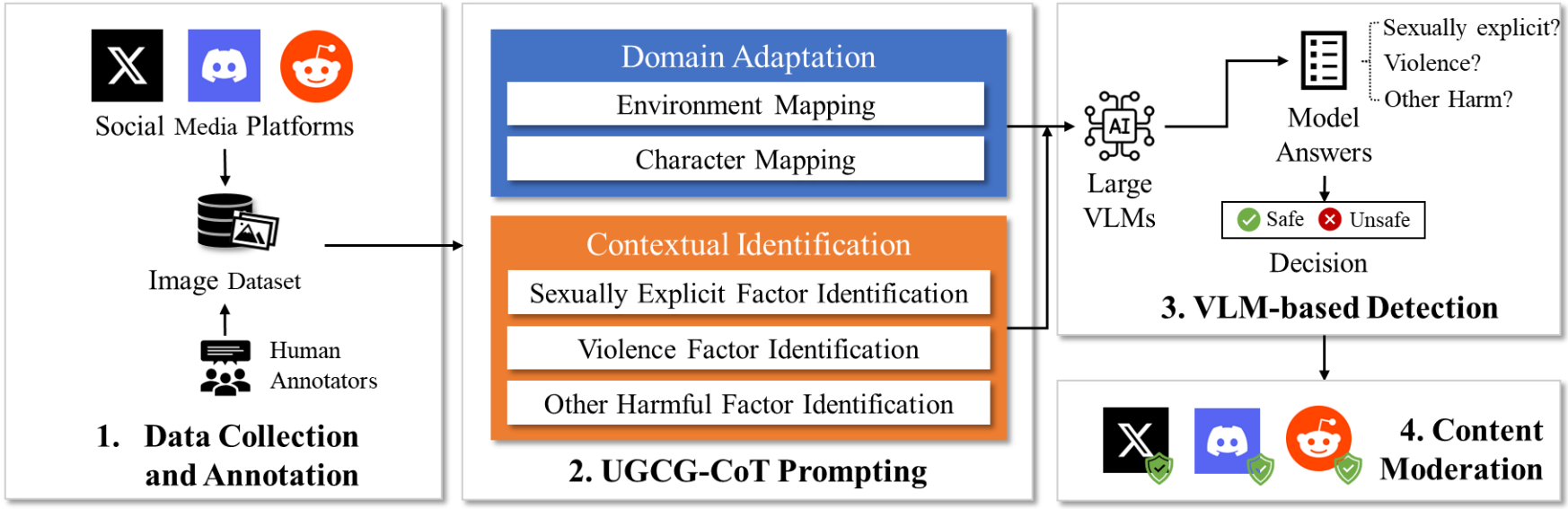

Sexually-explicit-UGCG

Image Type	State-of-the-Art Unsafe Image Detectors				
	Clarify	Yahoo Open NSFW	Amazon Rekog-nition	Micro-soft Azure	Google Vision AI
Sexually-explicit-human	88%	92%	98%	92%	98%
Sexually-explicit-anime	89%	81%	91%	90%	99%
Sexually-explicit-UGCG	13%	13%	17%	15%	67%

Utilizing LVLM against unsafe UGCG images

- UGCG-Guard



Other ongoing research

News

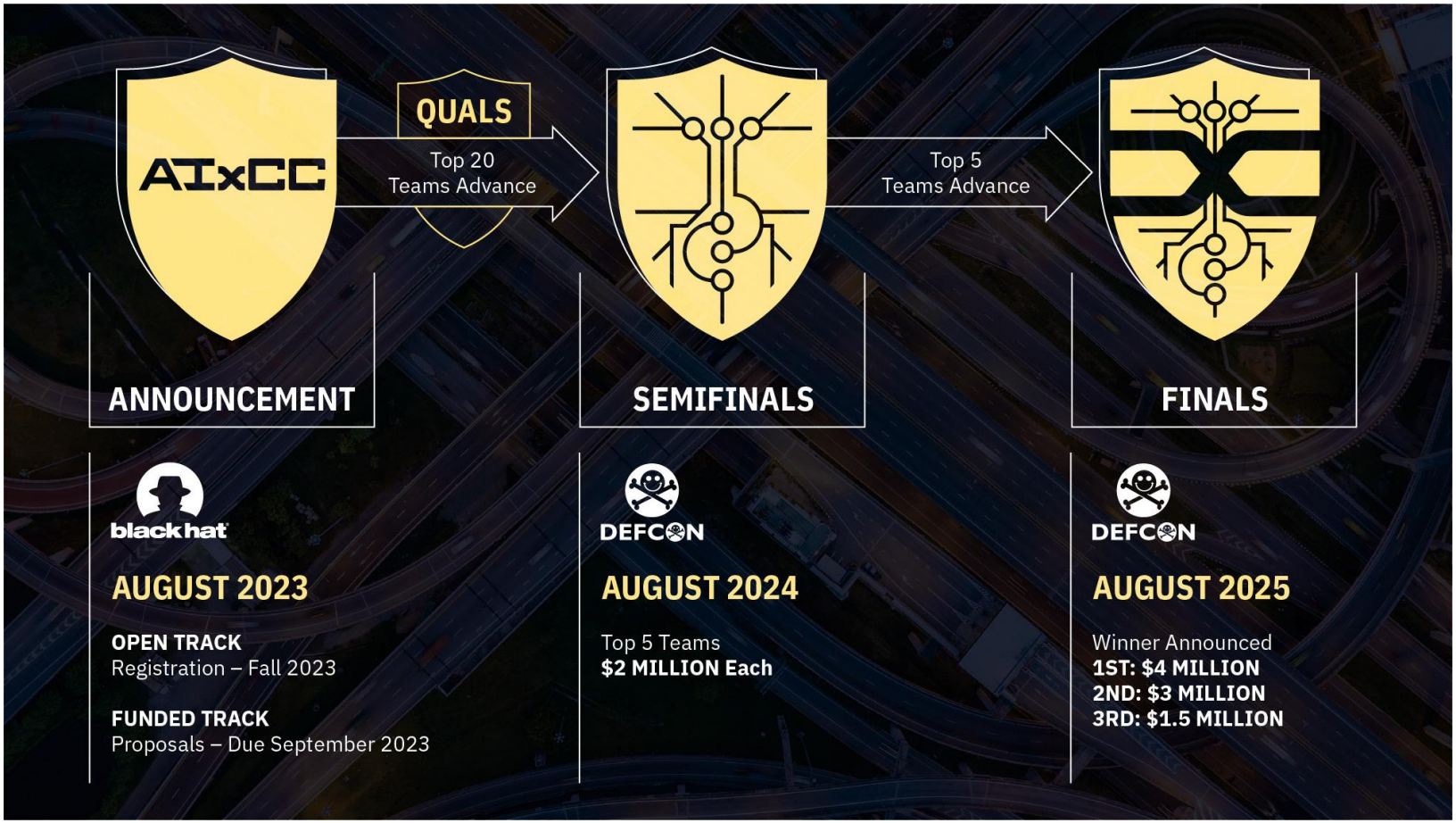
- 01/2025 [Paper]: Our paper "APPATCH: Automated Adaptive Prompting Large Language Models for Real-World Software Vulnerability Patching" has been accepted to **USENIX Security 2025**.
- 01/2025 [Paper]: Our paper "JBShield: Defending Large Language Models from Jailbreak Attacks through Activated Concept Analysis and Manipulation" has been accepted to **USENIX Security 2025**.

Ongoing Research for LLM Security

- **LLMs for Security/Safety**
  - Harmful Meme Detection/Hateful Video Detection
  - LLMs for Network Security
    - 5G Testing
    - Network Intrusion Detection
  - LLMs for Software Security
    - Software Vulnerability Detection/Repair
- **LLM Security/Safety**
  - Code Generation Model Hardening
  - Jailbreak Defense
  - …

DARPA's Artificial Intelligence Cyber Challenge (AIxCC)

- https://aicyberchallenge.com/



Emerging Domains

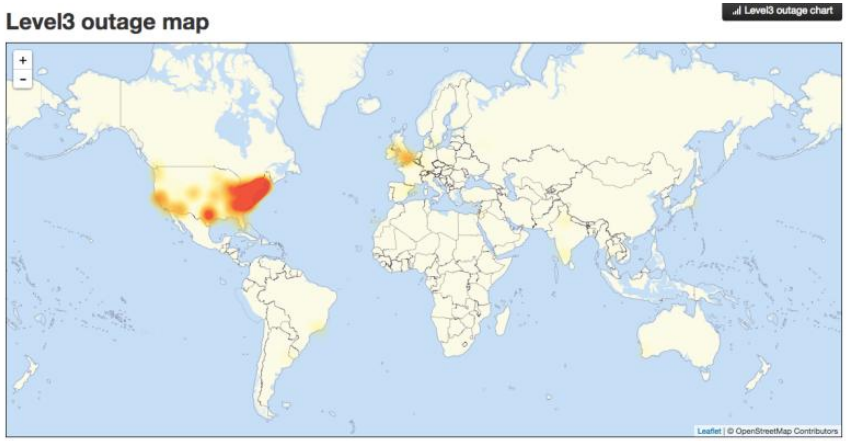
Mirai Botnet DDoS Attack

Someone is Using Mirai Botnet to Shut Down Internet for an Entire Country

Thursday, November 03, 2016 Swati Khandelwal

Note — We have published an updated article on what really happened behind the alleged DDoS attack against Liberia using Mirai botnet.

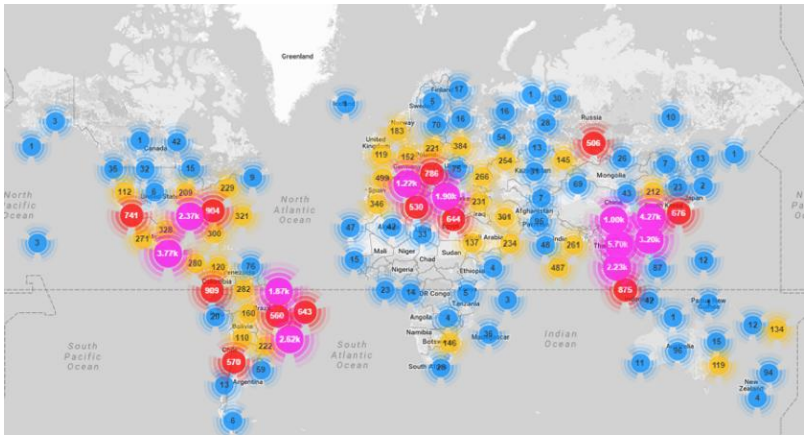
Someone is trying to take down the whole Internet of a country, and partially succeeded, by launching massive distributed denial-of-service (DDoS) attacks using a botnet of insecure IoT devices infected by the Mirai malware.



list of usernames and passwords included in the Mirai source code.

How Does Mirai work

- Mirai works by exploiting the weak security on many IoT devices, including routers, digital video records (DVRs), and webcams/security cameras.
- It operates by continuously scanning for IoT devices that are accessible over the internet and are protected by factory default or hardcoded user names and passwords.



Locations of all Mirai-infected devices

Username/Password	Manufacturer	Link to supporting evidence
admin/123456	ACTi IP Camera	https://ipvm.com/reports/ip-cameras-default-passwords-directory
root/anko	ANKO Products DVR	http://www.cctvforum.com/viewtopic.php?f=3&t=44250
root/pass	Axis IP Camera, et. al	http://www.cleancss.com/router-default/Axis/0543-001
root/vizxv	Dahua Camera	http://www.cam-it.org/index.php?topic=5192.0
root/888888	Dahua DVR	http://www.cam-it.org/index.php?topic=5035.0
root/666666	Dahua DVR	http://www.cam-it.org/index.php?topic=5035.0
root/7ujMko0vizxv	Dahua IP Camera	http://www.cam-it.org/index.php?topic=9396.0
root/7ujMko0admin	Dahua IP Camera	http://www.cam-it.org/index.php?topic=9396.0
666666/666666	Dahua IP Camera	http://www.cleancss.com/router-default/Dahua/DH-IPC-HDW4300C
root/dreambox	Dreambox TV receiver	https://www.satellites.co.uk/forums/threads/reset-root-password-plugin.101146/
root/zlxx	EV ZLX Two-way Speaker?	?
root/juantech	Guangzhou Juan Optical	https://news.ycombinator.com/item?id=11114012
root/xc3511	H.264 - Chinese DVR	http://www.cctvforum.com/viewtopic.php?f=56&t=34930&start=15
root/hi3518	HiSilicon IP Camera	https://acassis.wordpress.com/2014/08/10/i-got-a-new-hi3518-ip-camera-module-s/
root/klv123	HiSilicon IP Camera	https://gist.github.com/gabonator/74cdd6ab4f733ff047356198c781f27d
root/klv1234	HiSilicon IP Camera	https://gist.github.com/gabonator/74cdd6ab4f733ff047356198c781f27d
root/jvbzd	HiSilicon IP Camera	https://gist.github.com/gabonator/74cdd6ab4f733ff047356198c781f27d
root/admin	IPX-DDK Network Camera	http://www.ipxinc.com/products/cameras-and-video-servers/network-cameras/
root/system	IQinVision Cameras, et. al	https://ipvm.com/reports/ip-cameras-default-passwords-directory
admin/meinsm	Mobotix Network Camera	http://www.forum.use-ip.co.uk/threads/mobotix-default-password.76/
root/54321	Packet8 VOIP Phone, et. al	http://webcache.googleusercontent.com/search?q=cache:W1phozQZURUJ:community.freepbx.org/t/packet8-atas-phones/411
root/00000000	Panasonic Printer	https://www.experts-exchange.com/questions/26194395/Default-User-Password-for-Panasonic-DP-C405-Web-Interface.html
root/realtek	RealTek Routers	
admin/1111111	Samsung IP Camera	https://ipvm.com/reports/ip-cameras-default-passwords-directory
root/xmhdipc	Shenzhen Anran Security Camera	https://www.amazon.com/MegaPixel-Wireless-Network-Surveillance-Camera/product-reviews/B00EB6FNDI
admin/smcadmin	SMC Routers	http://www.cleancss.com/router-default/SMC/ROUTER
root/ikwb	Toshiba Network Camera	http://faq.surveillixdvrsupport.com/index.php?action=artikel&cat=4&id=8&artlang=en
ubnt/ubnt	Ubiquiti AirOS Router	http://setuprouter.com/router/ubiquiti/airos-airgrid-m5hp/login.htm
supervisor/supervisor	VideoIQ	https://ipvm.com/reports/ip-cameras-default-passwords-directory
root/<none>	Vivotek IP Camera	https://ipvm.com/reports/ip-cameras-default-passwords-directory
admin/1111	Xerox printers, et. al	https://atyourservice.blogs.xerox.com/2012/08/28/logging-in-as-system-administrator-on-your-xerox-printer/
root/Zte521	ZTE Router	http://www.ironbugs.com/2016/02/hack-and-patch-your-zte-f660-routers.html

Mirai Botnet Password Table

Car Hacking

Oct 20, 2014 | Julie Sinclair | 1 Comment |

Share 0 +1 0 Pin it

Is the Next Big Security Threat Car Hacking?

Today, automobiles offer you the convenience of hands free communication and GPS systems. As the automobile revolution advances and the cars

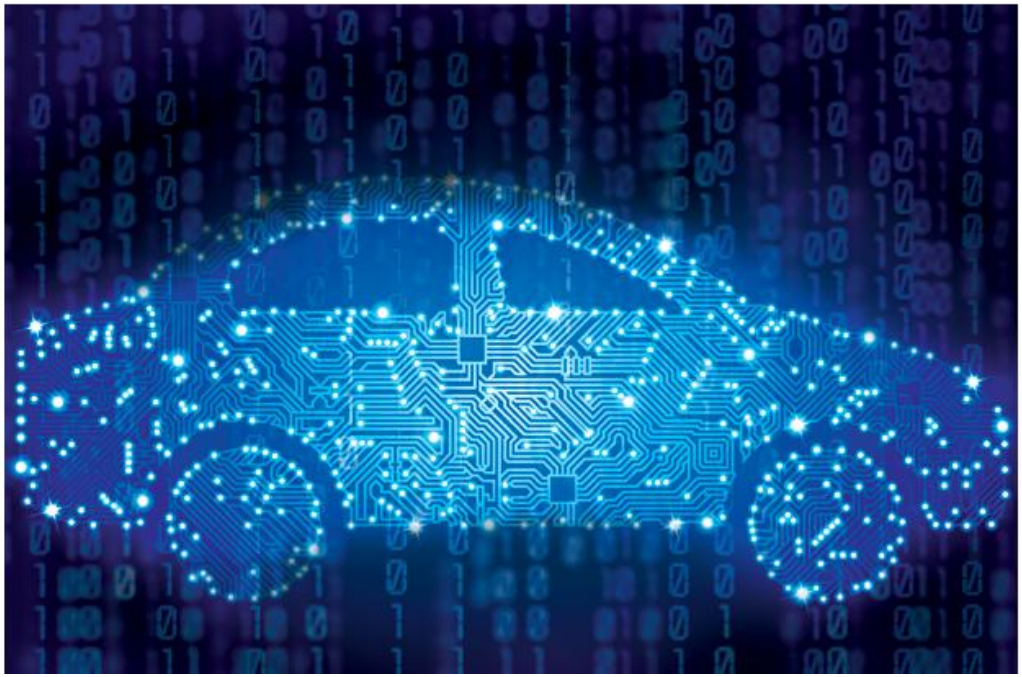
 on systems, they

Black Hat 2014: Hacking the Smart Car

By Mark Anderson

Posted 6 Aug 2014 | 16:00 GMT

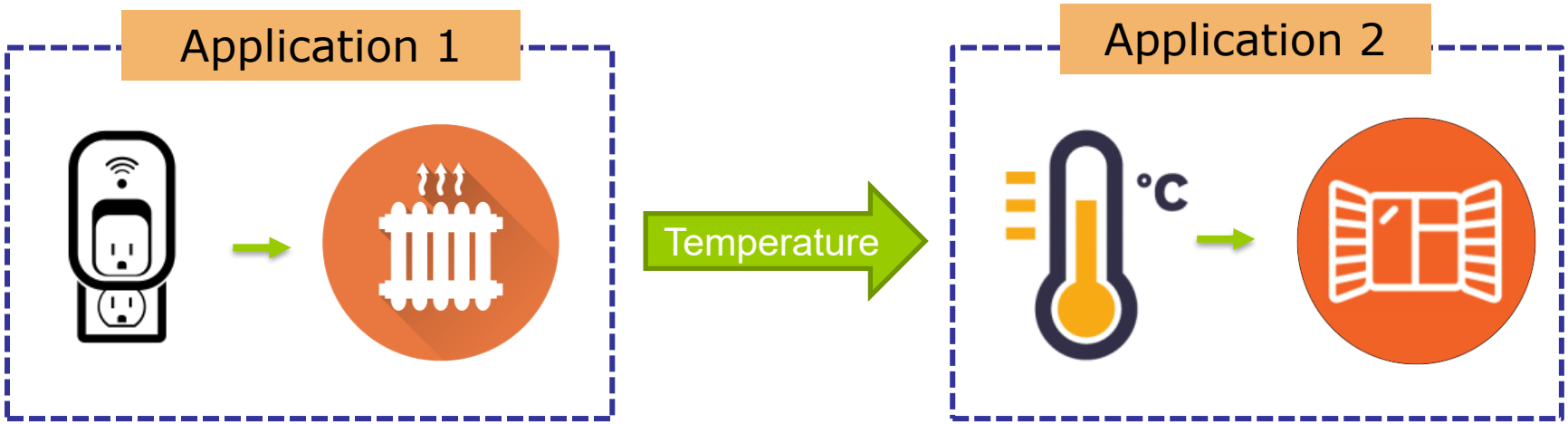
Share | Email | Print



Defcom'13: http://www.youtube.com/watch?v=n70hIu9lcYo
Black Hat'14: https://www.youtube.com/watch?v=rN58eb2XAb4
How to Hack a Tesla Model S (Defcom'15):
https://www.youtube.com/watch?v=KX_0c9R4Fng
Car Hacking Research: Remote Attack Tesla Motors by Keen Security Lab (2016):
https://www.youtube.com/watch?v=c1XyhReNcHY

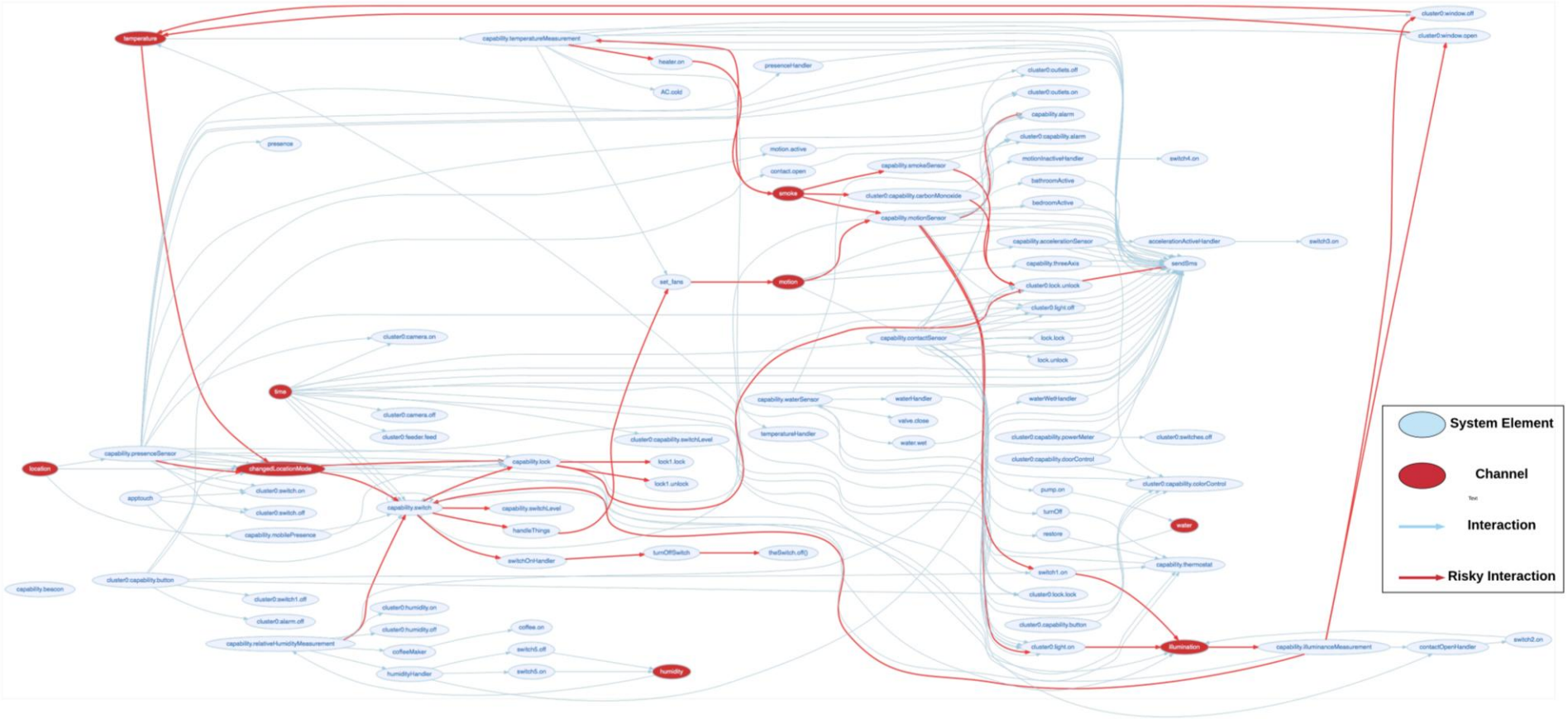
IoT Security - Physical Interaction Control

- Insecure interactions on physical channels in IoT devices



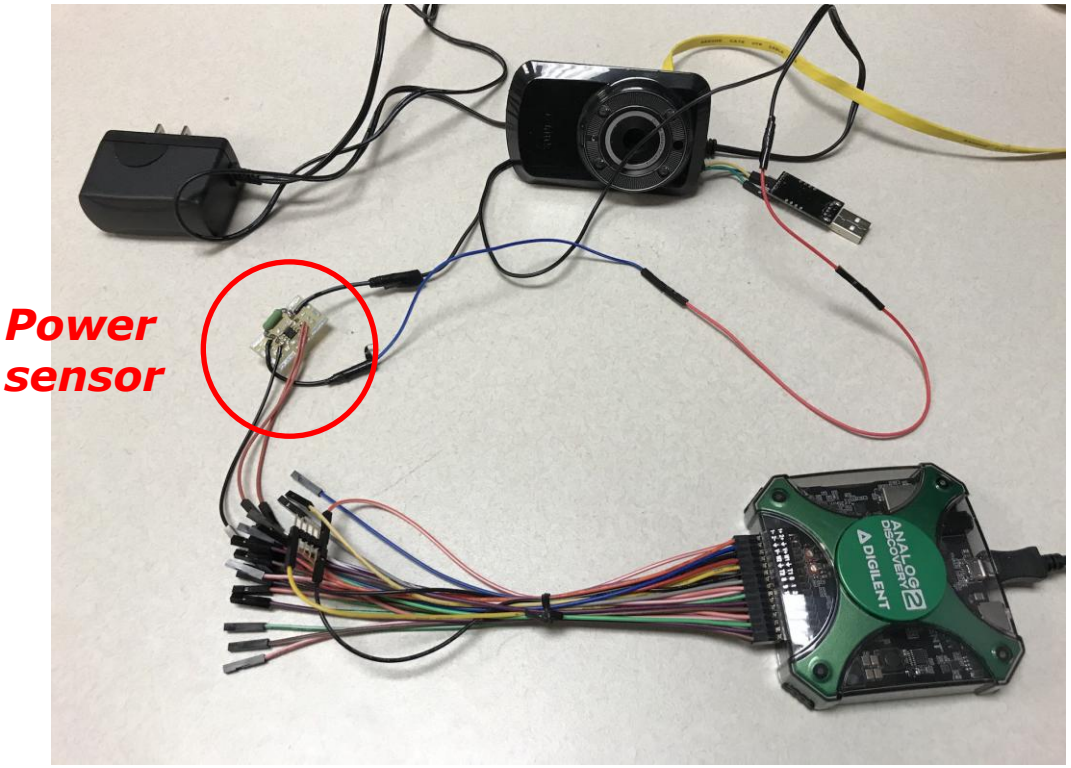
IoT Security - Physical Interaction Control

- 185 official SmartThings applications - 326 interaction chains
- 9 physical channels - 136 new interaction chains,
  52 high risky interaction chains



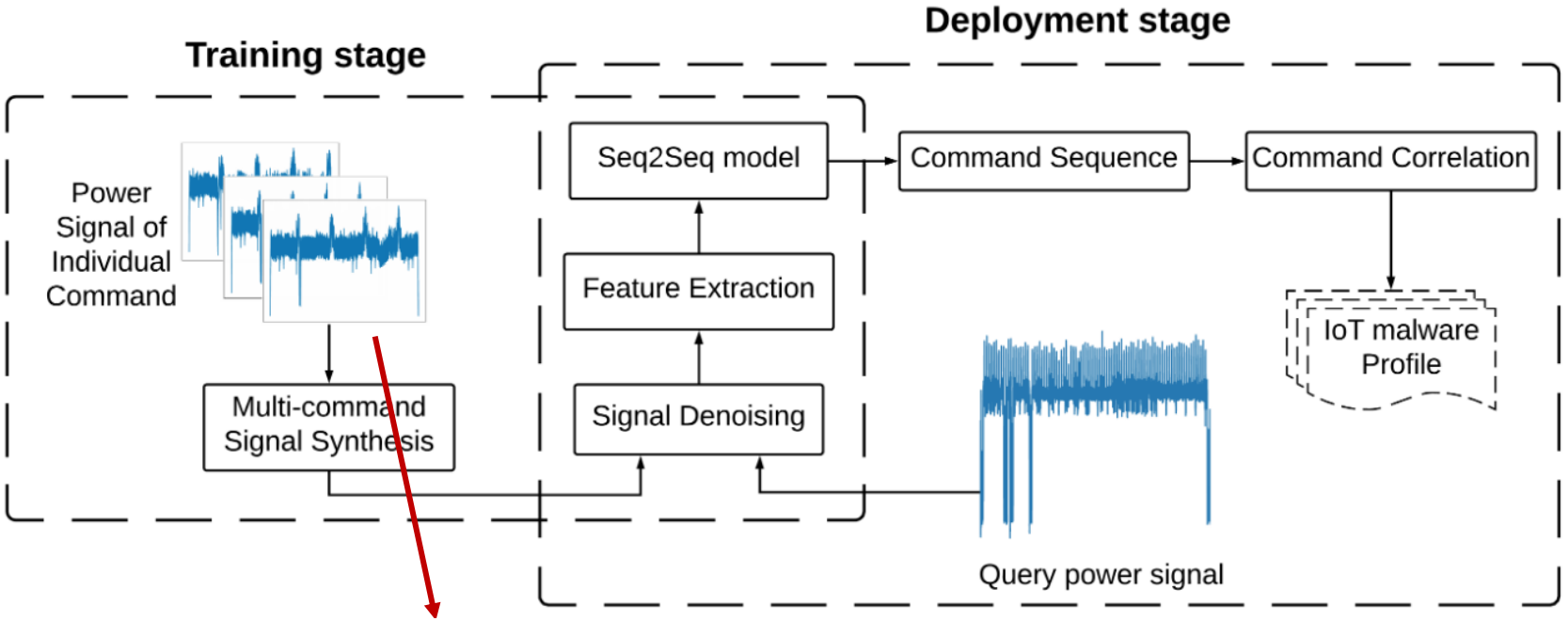
IoT Security – Malware Detection

- Early detection of IoT malware using side-channel power analysis



IoT Security – Malware Detection

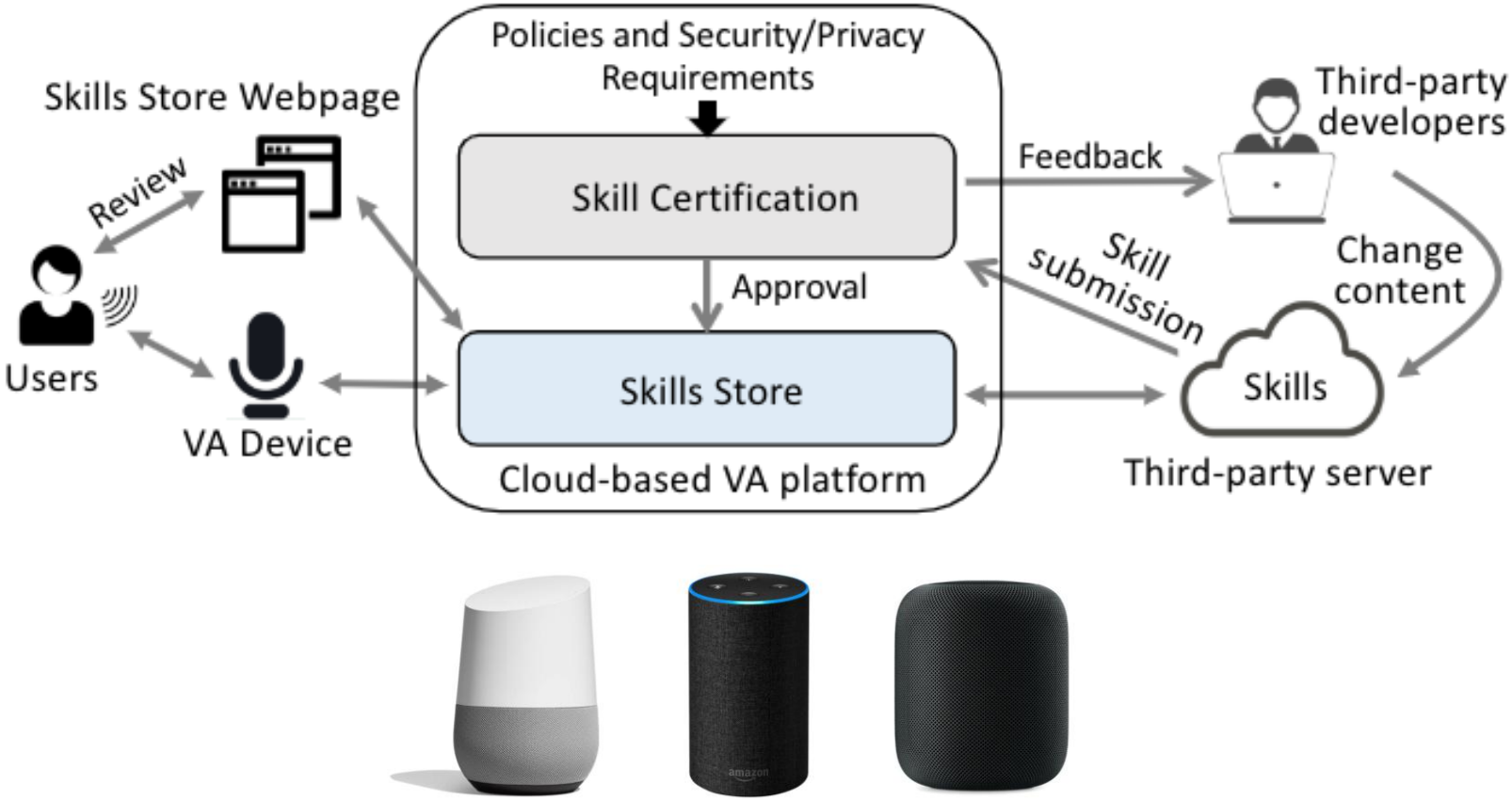
- Approach



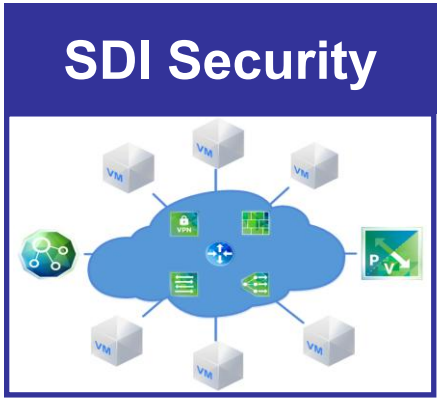
	CAT	CD	EXECUTE	ECHO	PS	GREP	LOGIN	KILL	WGET	METADATA	UNKNOWN
CAT	**73.86%**	0.59%	4.55%	1.19%	0.79%	3.17%	2.38%	0.79%	5.15%	5.94%	1.58%
CD	2.27%	**85.45%**	3.18%	5.45%	0.	0.	0.	2.73%	0.45%	0.	0.45%
EXECUTE	3.26%	0.65%	**76.02%**	1.14%	0.16%	2.45%	6.36%	2.28%	0.98%	5.22%	1.47%
ECHO	1.31%	3.93%	6.11%	**72.93%**	0.	3.06%	1.75%	9.17%	0.44%	0.87%	0.44%
PS	1.00%	0.	0.50%	0.10%	**82.5%**	9.90%	2.00%	0.10%	1.60%	1.20%	1.10%
GREP	2.50%	0.40%	3.10%	0.90%	14.30%	**67.30%**	4.90%	0.40%	4.00%	1.20%	1.00%
LOGIN	0.	0.	2.00%	0.40%	0.90%	4.30%	**83.90%**	0.30%	2.30%	2.70%	3.20%
KILL	2.47%	2.06%	6.58%	12.76%	0.	2.06%	1.65%	**66.67%**	3.29%	2.06%	0.41%
WGET	3.29%	0.09%	1.46%	0.	2.56%	3.20%	4.47%	1.19%	**74.98%**	6.39%	2.37%
METADATA	4.21%	0.34%	8.08%	0.46%	0.34%	1.37%	6.71%	1.02%	2.28%	**72.92%**	2.28%
UNKNOWN	1.59%	0.	0.40%	1.39%	0.40%	0.20%	3.37%	0.20%	1.79%	2.78%	**87.90%**

IoT Security – Voice Assistant (VA)

- Analyze and integrate **Voice Assistant (VA)** platforms

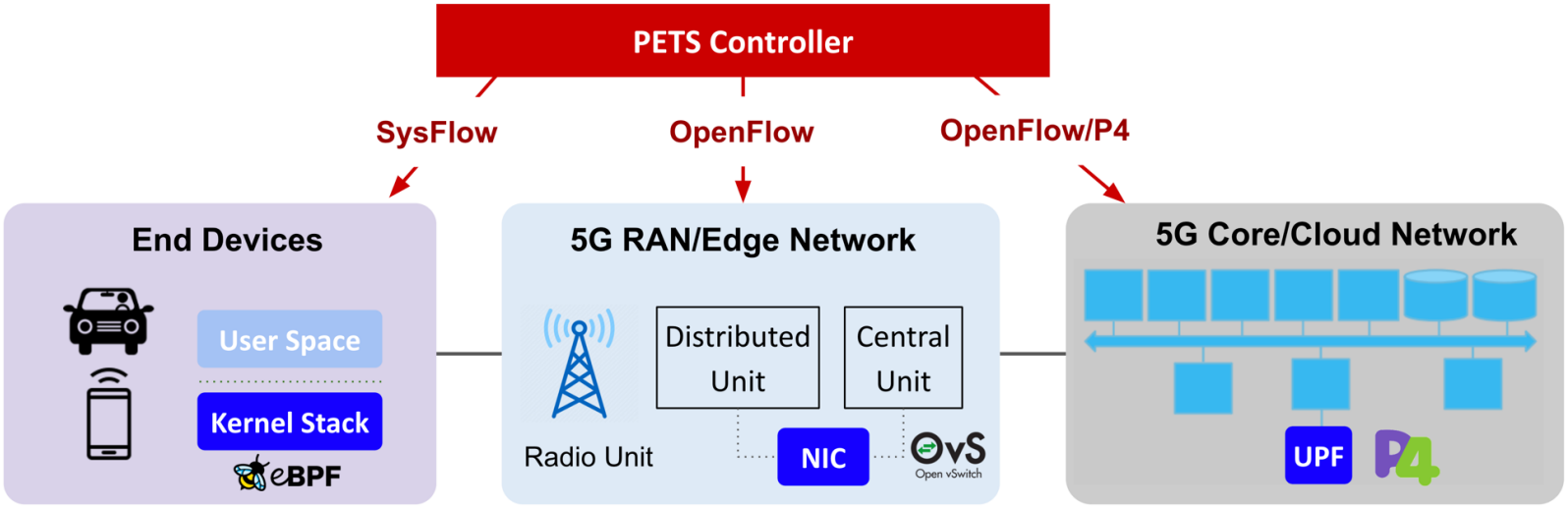


Emerging Domains



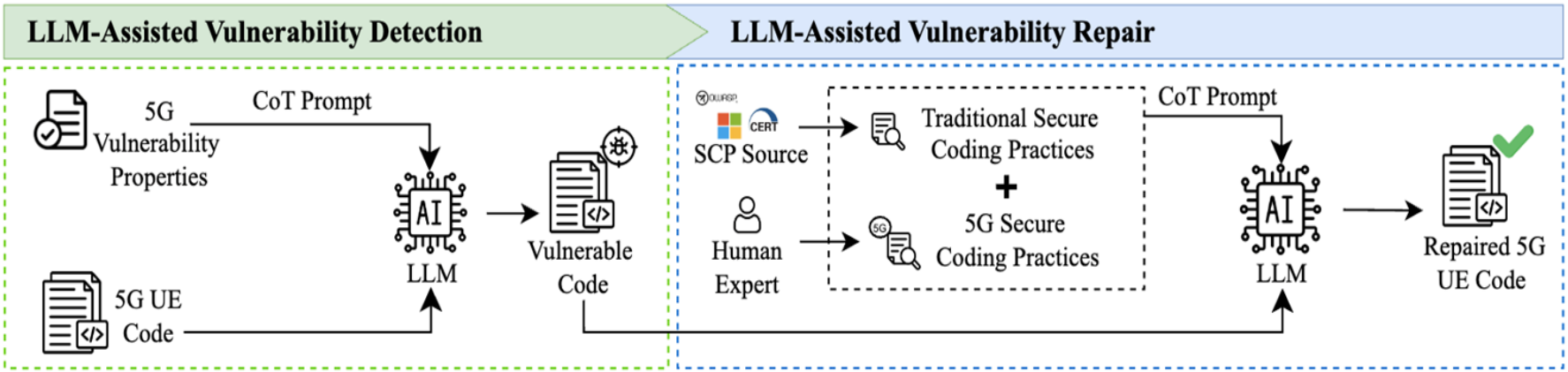
5G/Next-G Security

- PETS: Programmable Zero-Trust Security for Operating Through 5G Infrastructure



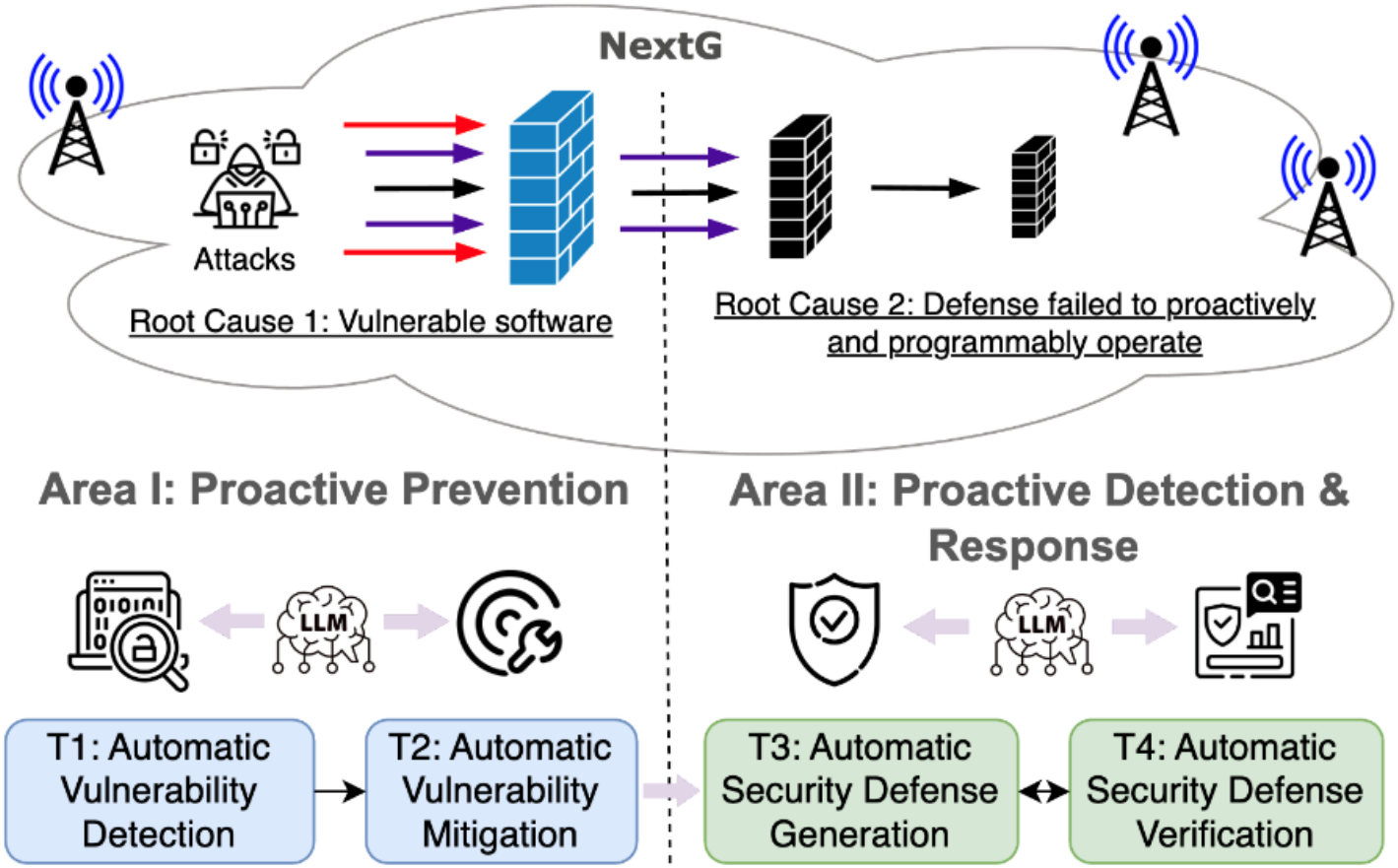
5G/Next-G Security

- Towards **LLM-Assisted Vulnerability Detection and Repair** for Open-Source 5G UE Implementations (FutureG 2025)



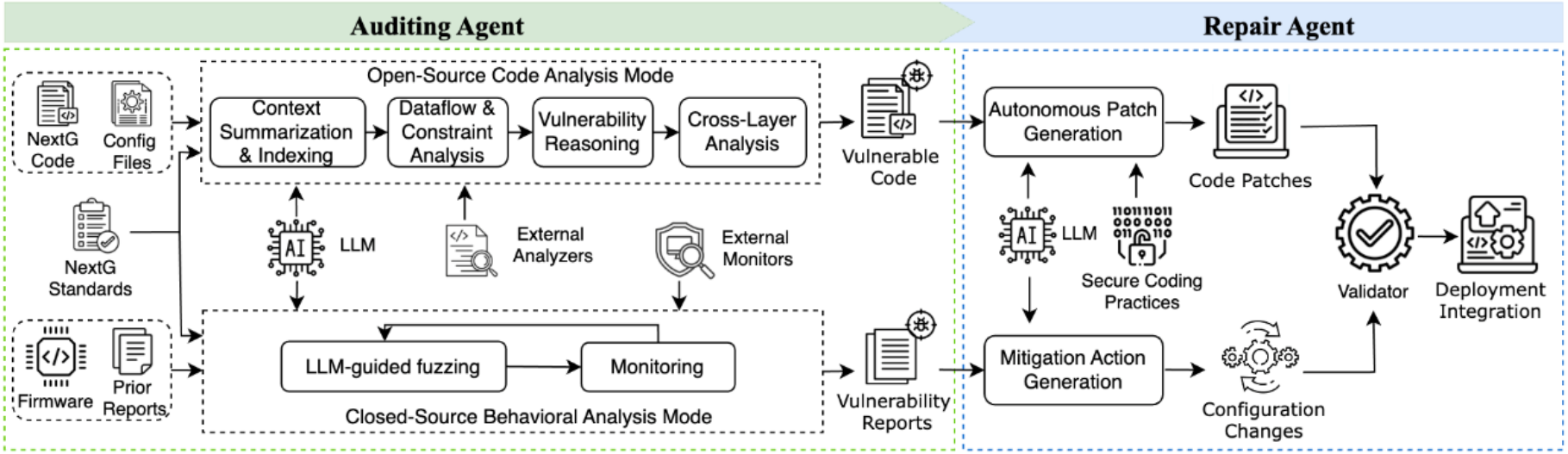
5G/Next-G Security

- LLM-Agentic Framework for NextG Vulnerability Detection and Mitigation



5G/Next-G Security

- LLM-powered Agentic and Proactive Security Defense for NextG Network Systems

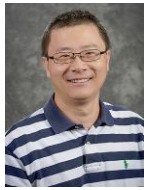


S²OS

- S²OS: Enabling Infrastructure-wide Programmable Security with SDI ($3M, NSF/VMware SDI-CSCS)

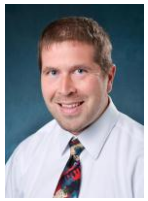
S²OS Team

- **Leading experts on security, networking, and systems**



- Guofei Gu, Texas A&M University
  - Network **Security** and **SDN Security**

- Hongxin Hu, Clemson University
  - **NFV** and **SDN Security**

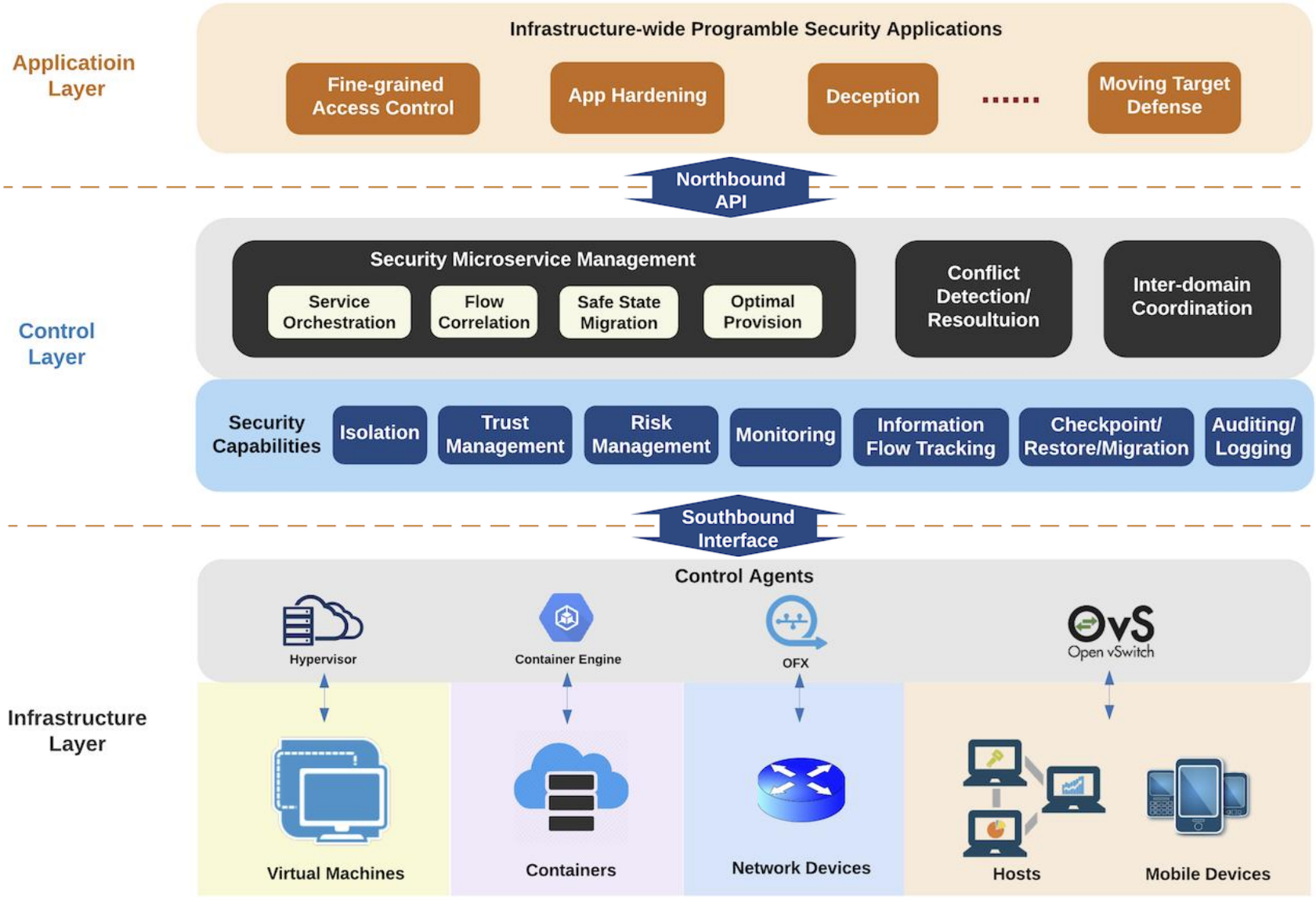


- Eric Keller, University of Colorado, Boulder
  - **Networking, Virtualization** and **SDN**

- Zhiqiang Lin, Ohio State University
  - **Virtualization** and **Systems Security**

- Donald Porter, UNC Chapel Hill
  - **Operating Systems**, **Virtualization**, **Storage**, and **Systems Security**

S²OS Overview: Layered OS Design

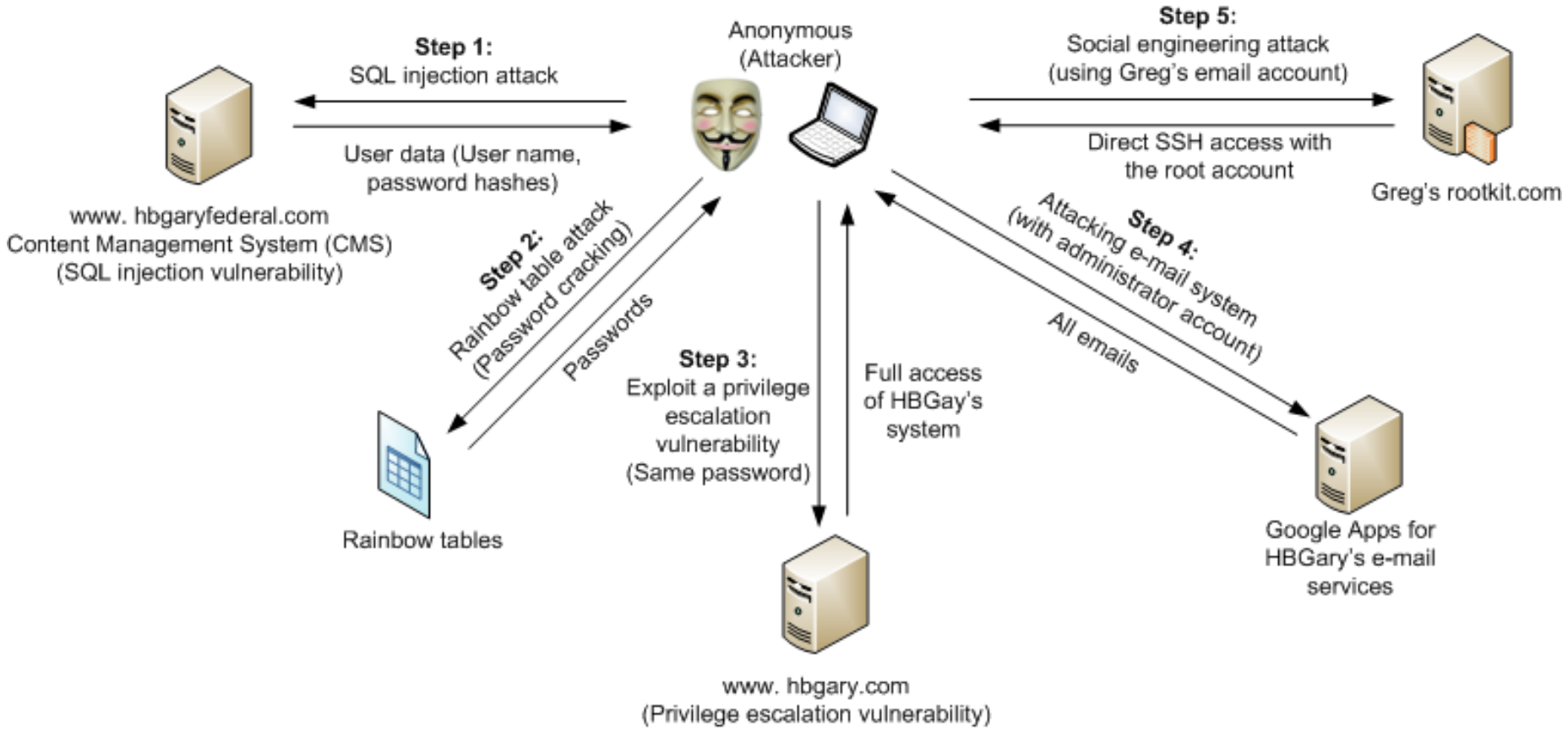


HBGary Hack

- **Anonymous speaks: the inside story of the HBGary hack**
  - http://arstechnica.com/tech-policy/2011/02/anonymous-speaks-the-inside-story-of-the-hbgary-hack/
- **Back ground**
  - HBGary Federal is a security consultant company
  - Anonymous is the well-known hacker group that was responsible for the attacks on Amazon, PayPal, MasterCard, Visa and the Swiss bank PostFinance recently.

- **Story**
  - The CEO of HBGary Federal, Aaron Barr, claimed he knew the true identities of the leaders behind Anonymous, and threatened to reveal them at the 2011 RSA security conference. Anonymous hacked and defaced the HBGary Web site, and compromised its servers. Tens of thousands of HBGary e-mails were then exposed on the Web.

HBGary Hack



- What can we learn from this attack?

HBGary Hack – Lessons Learned

- **Use Off-the-Shelf Software/**
  - Is off-the-shelf software more secure than custom-made solutions?
- **Patch Your Systems Regularly**
- **Store Passwords Securely**
  - HBGary used MD5 to hash the passwords, it used MD5 badly: there was no iterative hashing and no salting
- **Force Users to Create Complex Passwords**
- **Don't Reuse Passwords**
- **Delete Sensitive E-mails**
  - An e-mail containing the root password for the rootkit.com site
- **Educate Against Social Engineering**
  - Anonymous had a root password, but they couldn't access the rootkit.com server because it didn't allow root access from outside of the firewall--a wise security move. To get this, they used Greg Hoglund's e-mail account to make contact with somebody who had root access to the server

Questions?