# Developing Robust Models, Algorithms, Databases and Tools with Applications to Cybersecurity and Healthcare

## ML PhD Dissertation Defense


Scott Freitas
ML PhD Candidate

Committee

Diyi Yang

Srijan Kumar

B. Aditya Prakash

Hanghang Tong

Polo Chau

# Machine learning is all around us

Cybersecurity → Catch bad guys

Autonomous Vehicles

Face unlock

Traffic & commute forecast

Email spam filters

Infrastructure monitoring

Search results

News recommendations

Social media newsfeeds

Mobile computational photography

Voice assistants

Voice dictation

Health informatics

Personalized music & radio

Entertainment recommendations

Spelling & text prediction

# Machine learning is all around us

Cybersecurity → Catch bad guys

Autonomous Vehicles → Less accidents

Face unlock

Traffic & commute forecast

Email spam filters

Infrastructure monitoring

Search results

News recommendations

Social media newsfeeds

Mobile computational photography

Voice assistants

Voice dictation

Health informatics

Personalized music & radio

Entertainment recommendations

Spelling & text prediction

# Machine learning is all around us

Cybersecurity → Catch bad guys

Autonomous Vehicles → Less accidents

Face unlock

Traffic & commute forecast

Email spam filters

Infrastructure monitoring → Prevent outages

Search results

News recommendations

Social media newsfeeds

Mobile computational photography

Voice assistants

Voice dictation

Health informatics

Personalized music & radio

Entertainment recommendations

Spelling & text prediction

# Machine learning is all around us

Cybersecurity → Catch bad guys

Autonomous Vehicles → Less accidents

Face unlock

Traffic & commute forecast

Email spam filters

Infrastructure monitoring → Prevent outages

Search results

News recommendations

Social media newsfeeds → Connect people

Mobile computational photography

Voice assistants

Voice dictation

Health informatics

Personalized music & radio

Entertainment recommendations

Spelling & text prediction

# Machine learning is all around us

Cybersecurity → Catch bad guys
Autonomous Vehicles → Less accidents
Face unlock
Traffic & commute forecast
Email spam filters
Infrastructure monitoring → Prevent outages
Search results
News recommendations
Social media newsfeeds → Connect people
Mobile computational photography
Voice assistants
Voice dictation
Health informatics → Life saving
Personalized music & radio
Entertainment recommendations
Spelling & text prediction

# Machine learning is all around us

Cybersecurity → Catch bad guys
Autonomous Vehicles → Less accidents
Face unlock
Traffic & commute forecast
Email spam filters
Infrastructure monitoring → Prevent outages
Search results
News recommendations
Social media newsfeeds → Connect people
Mobile computational photography
Voice assistants
Voice dictation
Health informatics → Life saving
Personalized music & radio
Entertainment recommendations
Spelling & text prediction → Send right message

# Machine learning is all around us

| | |
|---|---|
| Cybersecurity | Catch bad guys |
| Autonomous Vehicles | Less accidents |
| Face unlock | Better privacy |
| Traffic & commute forecast | On-time |
| Email spam filters | Filter dangerous emails |
| Infrastructure monitori | Prevent outages |
| Search results | Improved information |
| News recommendatio | Better stories |
| Social media newsfeed | Connect people |
| Mobile computational | Create memories |
| Voice assistants | Help non-native users |
| Voice dictation | Record messages |
| Health informatics | Life saving |
| Personalized music & radio | Improve mood |
| Entertainment recommendations | Save time |
| Spelling & text prediction | Send right message |

# Why Robust Machine Learning?


It’s disturbingly easy to trick AI into doing something deadly Vox

How “adversarial attacks” can mess with self-driving cars, medicine, and the military.

By Sigal Samuel | Apr 8, 2019, 9:10am EDT

UHS Ransomware Attack Cost $67M in Lost Revenue, Recovery Efforts

The ransomware attack that struck all 400 UHS care sites and caused three weeks of EHR downtime in September, cost the health system $67 million in recovery costs and lost revenue.

Security News This Week: An Unprecedented Cyberattack Hit US Power Utilities WIRED

Exposed Facebook phone numbers, an XKCD breach, and more of the week's top security news.

## Detect and prevent attacks on

- critical infrastructure
- self driving cars
- enterprise networks

## Improve decision making

- robust to noise
- identify weak points
- quantify vulnerability


Microsoft ATP


IBM Research


amazon


NAWC
Weapons Division
NAVAL AIR WARFARE CENTER


**Machine learning** is transforming the public and private sector.

How do we **protect it?**

# Why do we need robust ML?

Cybersecurity → Backdoors

Autonomous Vehicles → Safe driving

Face unlock

Traffic & commute forecast

Email spam filters

Infrastructure monitoring → Prevent outages

Search results

News recommendations

Social media newsfeeds → Pigeonhole

Mobile computational photography

Voice assistants

Voice dictation

Health informatics → Save lives

Personalized music & radio

Entertainment recommendations

Spelling & text prediction

# Robust machine learning protects us

Cybersecurity → Backdoors
Autonomous Vehicles → Safe driving
Face unlock → Annoying
Traffic & commute forecast → Late
Email spam filters → Missed important email
Infrastructure monitori → Prevent outages
Search results → Spread misinformation
News recommendatio → Fake news
Social media newsfeed → Pigeonhole
Mobile computational → Miss memories
Voice assistants → Alienate users
Voice dictation → Wrong messages
Health informatics → Save lives
Personalized music & radio → Ruin mood
Entertainment recommendations → Waste time
Spelling & text prediction → Send wrong message

# Dissertation Research Mission

Address large-scale societal problems in cybersecurity and healthcare through **the lens of robust machine learning**

Part I:
## Tools

**Robustness Survey** Summarize robustness literature TKDE 2021 (under review)

**TIGER** Vulnerability and robustness toolbox CIKM 2021

Part II:
## Algorithms

**D²M** Quantify network robustness + mitigate attacks SDM 2020

Part III:
## Databases

**MalNet-Graph** Largest cybersecurity graph database NeurIPS 2021

**MalNet-Image** Largest cybersecurity image database Submitting to CIKM 2022

Part IV:
## Models

**UnMask** Identify robust features in images IEEE Big Data 2020

**REST** Identify robust signals in health data Web Conference 2020

# Dissertation Research Mission

Address large-scale societal problems in cybersecurity and healthcare through **the lens of robust machine learning**

Part I:
**Tools**

**Robustness Survey** Summarize robustness literature TKDE 2021 (under review)

**TIGER** Vulnerability and robustness toolbox CIKM 2021

Part II:
**Algorithms**

**D²M** Quantify network robustness + mitigate attacks SDM 2020

Part III:
**Databases**

**MalNet-Graph** Largest cybersecurity graph database NeurIPS 2021

**MalNet-Image** Largest cybersecurity image database Submitting to CIKM 2022

Part IV:
**Models**

**UnMask** Identify robust features in images IEEE Big Data 2020

**REST** Identify robust signals in health data Web Conference 2020

# Part I: Why do we Need Robust **Tools**?

To **democratize knowledge**, and **equip users** to develop robust ML systems

- Robustness knowledge is currently scattered across disparate fields
- Key data and libraries are in the possession of a few industry labs

# Robustness Survey

## Graph Vulnerability and Robustness: A Survey

TKDE 2021 (under review)


Scott Freitas
Georgia Tech

Diyi Yang
Georgia Tech

Srijan Kumar
Georgia Tech

Hanghang Tong
UIUC

Polo Chau
Georgia Tech

# Survey Highlights

Comprehensively survey and compare **85** high-impact and recent papers in the field of graph robustness

Each row is one work, columns are grouped into one of three categories—robustness measures, attacks, and defenses.

| Work | 2.1.1 Binary Connectivity | 2.1.2 Vertex Connectivity | 2.1.3 Edge Connectivity | 2.1.4 Diameter | 2.1.5 Average Distance | 2.1.6 Avg. Vertex Betweenness | 2.1.7 Avg. Edge Betweenness | 2.1.8 Global Clustering Coefficient | 2.1.9 Largest Connected Component | 2.2.1 Spectral Radius | 2.2.2 Spectral Gap | 2.2.3 Natural Connectivity | 2.2.4 Spectral Scaling | 2.2.5 Generalized Robustness Index | 2.3.1 Algebraic Connectivity | 2.3.2 Number of Spanning Trees | 2.3.3 Effective Resistance | 3.3 Node Removal | 3.3 Edge Removal | 4.1.1 Edge Addition | 4.1.2 Edge Rewiring | 4.1.3 Node Monitoring | Publication Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Robustness Measures | | | | | | | | | | | | | | | | | Attack | | Defense | | | Where |
| Albert,et.al. [12] | | | | ● | ● | | | | ● | | | | | | | | | ● | | | | | Nature |
| Alenazi,et.al. [13] | | | | | ● | ● | | ● | | | ● | ● | | | ● | | | ● | | ● | | | RNDM |
| Alenazi,et.al. [14] | | | | | ● | ● | | ● | | | | | ● | ● | ● | | ● | | | | | | DRCN |
| Baig,et.al. [15] | | | | | | | | | | | | | | | | | | ● | | | | | Web |
| Baras,et.al. [16] | | | | | | | | | | | | | | | ● | ● | ● | | | ● | | | CDC |
| Derdica [17] | | | ● | | | | | | | | | | | | | | | | | | | | A-POL |
| Bernstein,et.al. [18] | | | | | | | | | | | | | | | | | | | ● | | | | INFO |
| Beygelzimer,et.al. [3] | | | | ● | ● | | | | ● | | | | | | | | | ● | | ● | ● | | Physica A |
| Bigdeli,et.al. [19] | | | | | | ● | | | | | | | | | ● | | | | | | | | SIMPLEX |
| Bishop,et.al. [20] | | | | | | | | | | ● | | | | | | | | | ● | | | | EPL |
| Bocca,et.al. [21] | | | | ● | ● | ● | ● | ● | ● | | | | | | ● | | | ● | ● | | | | PR |
| Borgatti,et.al. [22] | | | | | | ● | | | | | | | | | | | | ● | ● | ● | | | SN |
| Briesemeis.,et.al. [23] | | | | | | | | | | | | | | | | | | | | | | ● | WORM |
| Buldyrev,et.al. [24] | | | | | | | | | ● | | | | | | | | | ● | | | | | Nature |
| Byrne,et.al. [25] | | ● | ● | | ● | | | | | | | | | | ● | | | | | | | | Sandia |
| Caballero,et.al. [26] | | | | | | | | | ● | | | | | | | | | | ● | | | | NDSS |
| Callaway,et.al. [27] | | | | | | | | | ● | | | | | | | | | ● | | | | | PRL |
| Chan,et.al. [28] | | | | | | | | | | | | ● | | | | | | ● | ● | | | | SDM |
| Chakrabarti,et.al. [29] | | | | | | | | | | ● | | | | | | | | | | | | ● | TOPS |
| Chan,et.al. [7] | | | | | | | | | | | ● | ● | ● | | ● | ● | ● | | | | ● | | DMKD |
| Chen,et.al. [30] | | | | | | | | | | ● | | | | | | | | | | | | ● | TKDE |
| Chen,et.al. [31] | | | | | | | | ● | | ● | | | | | | | | ● | | | | | ICDM |
| Chen,et.al. [32] | | | | | | | | | | ● | | | | | | | | | ● | ● | | | TKDD |
| Crucitti,et.al. [33] | | | | | ● | | | | | | | | | | | | | ● | | | | | PRE |
| Dekker [34] | | ● | | ● | | | | | | | | | | | | | | ● | | | | | ACSC |
| Derrible,et.al. [35] | | | | ● | ● | | | ● | | | | | | | | | | | | | | | Physica A |
| Duan,et.al. [36] | | | | | ● | ● | | ● | ● | | | | | | | | | ● | | | | | Physica A |
| Ellens,et.al. [37] | | | | | | | | | | | | | | | | | ● | | | ● | | | LAA |
| Ellens,et.al. [6] | ● | ● | ● | ● | ● | ● | ● | ● | | | | | | | ● | ● | ● | | | | | | arXiv |
| Estrada,et.al. [38] | | | | | | | | | | | ● | | ● | | | | | | | | | | Physica B |
| Estrada,et.al. [39] | | | | | | | | | | | ● | | ● | | | | | | | | | | EPL |
| Freitas,et.al. [40] | | | | | | | | | | | | | | | | | | | | | | ● | SDM |
| Freitas,et.al. [41] | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | arXiv |
| Gao,et.al. [42] | | | | | | | | | ● | | | | | | | | | ● | | | | | PRL |
| Ghosh,et.al. [43] | | | | | | | | | | | | | | | | | ● | | | | | | SIREV |
| Holme,et.al. [44] | | | | | ● | | | | ● | | | | | | | | | ● | ● | | | | PRE |
| Holmgren [45] | | | | | ● | | | ● | ● | | | | | | | | | ● | | | | | RA |
| Jamakovic,et.al. [46] | | ● | ● | | | | | | | | | | | | ● | | | ● | ● | | | | NGI |
| khalil,et.al. [47] | | | | | | | ● | | | ● | | | | | | | | | ● | ● | | | KDD |
| Kinney,et.al. [48] | | | | | ● | | | | | | | | | | | | | ● | | | | | EPJ B |
| Klau,et.al. [49] | ● | ● | | ● | ● | | | | ● | | | | | | | | | ● | ● | | | | Net. Anal. |
| Latora,et.al. [50] | | | | | ● | | | | | | | | | | | | | ● | ● | ● | | | PRE |
| Le,et.al. [51] | | | | | | | | | | | | | | | | | | | ● | | | | SDM |
| Leskovec,et.al. [52] | | | | | | | | | | | | | | | | | | | | | | ● | KDD |
| Liu,et.al. [53] | | | | | ● | ● | ● | | | | | ● | | | ● | | | ● | ● | | | | FCS |
| Lu,et.al. [54] | | | | | | ● | ● | | | | | | | | | | | ● | ● | | | | PLOS One |
| Malliaros,et.al. [55] | | | | | | | | | | | | | | ● | | | | | | | | | SDM |

# Survey Overview

## Robustness Measures

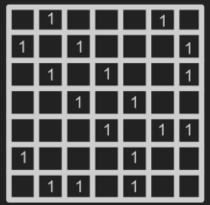
Summary & comparison of 18 **robustness metrics**

### Graph Measures

- Diameter
- Average distance
- Edge connectivity

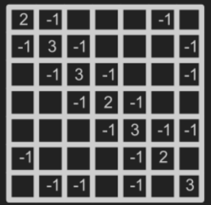
### Adjacency Measures

- Spectral radius
- Spectral scaling
- Natural connectivity



### Laplacian Measures

- Effective Resistance
- Algebraic connectivity
- Number of Trees



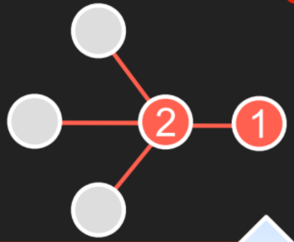
## Failure Scenarios

Study of **failure scenarios** on various **graph** types

### Natural Failure

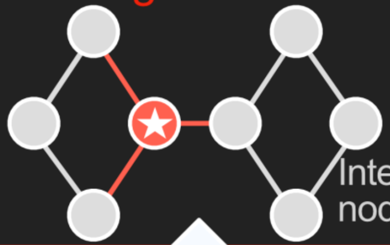
Failure of a single node

### Cascading Failure




Sequential failure of nodes

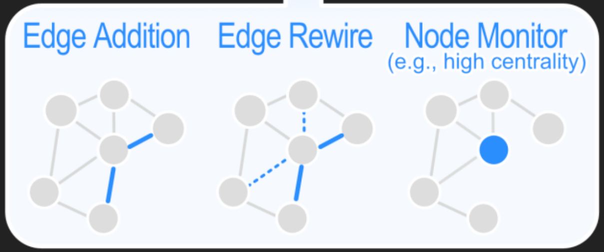
### Targeted Attack

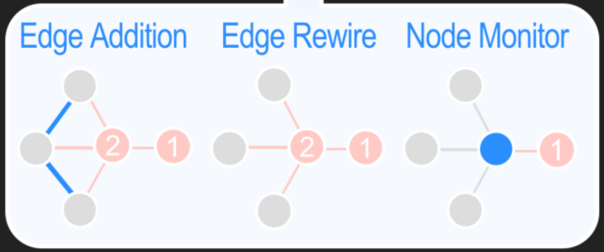


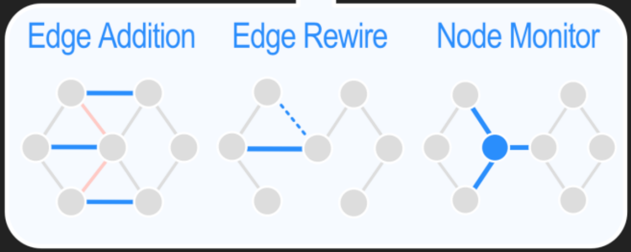
Intentional node damage

## Defense Techniques

Discussion of **defenses** against multiple **failures**



# Recap: Survey Contributions

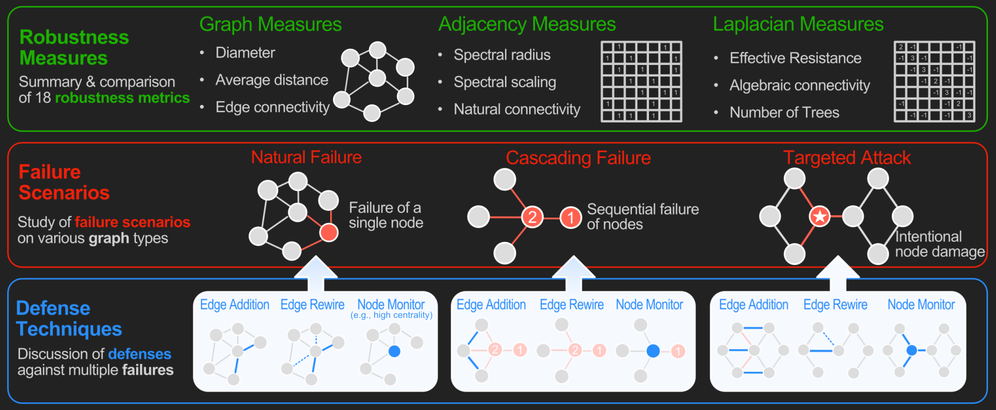
**C1.** Summary and comparison of 18 robustness measures

**C2.** Exploration of robustness measure applications

**C3.** Overview of network attack strategies

**C4.** Comparison of network defense mechanisms

**C5.** Highlight open problems and research directions



# TIGER Robustness Toolbox

## Evaluating Graph Vulnerability and Robustness using TIGER

CIKM 2021


Scott Freitas
Georgia Tech

Diyi Yang
Georgia Tech

Srijan Kumar
Georgia Tech

Hanghang Tong
UIUC

Polo Chau
Georgia Tech

# TIGER Overview

Available at https://github.com/safreita1/TIGER

Part of **Nvidia Data Science Teaching Kit**

**TIGER** is a GPU accelerated Python library and part of the Nvidia Data Science Teaching Kit

1. **Quantify** network *vulnerability* and *robustness*
2. **Simulate** a variety of network attacks, cascading failures and spread of dissemination of entities
3. **Augment** a network's structure to resist *attacks* and recover from *failure*
4. **Regulate** the dissemination of entities on a network (e.g., viruses, propaganda)

# TIGER Contributions

1. First open-source Python toolbox to evaluate graph vulnerability and robustness
2. **22 robustness measures** with original and fast approximate versions
3. **17 failure and attack mechanisms**
4. **15 defense techniques** (heuristic and optimization-based)
5. **4** network simulation techniques for cascading failures and entity dissemination

# Quantifying Robustness

| Robustness Measure | Category |
|---|---|
| Vertex connectivity | Graph |
| Edge connectivity | Graph |
| Diameter | Graph |
| Average distance | Graph |
| Average inverse distance | Graph |
| Average vertex betweenness | Graph |
| Average edge betweenness | Graph |
| Global clustering coefficient | Graph |
| Largest connected component | Graph |
| Spectral radius | Adjacency matrix |
| Spectral gap | Adjacency matrix |
| Natural connectivity | Adjacency matrix |
| Spectral scaling | Adjacency matrix |
| Generalized robustness index | Adjacency matrix |
| Algebraic connectivity | Laplacian matrix |
| Number of spanning trees | Laplacian matrix |
| Effective resistance | Laplacian matrix |

22 robustness measures

9 graph measures +
2 approximate versions

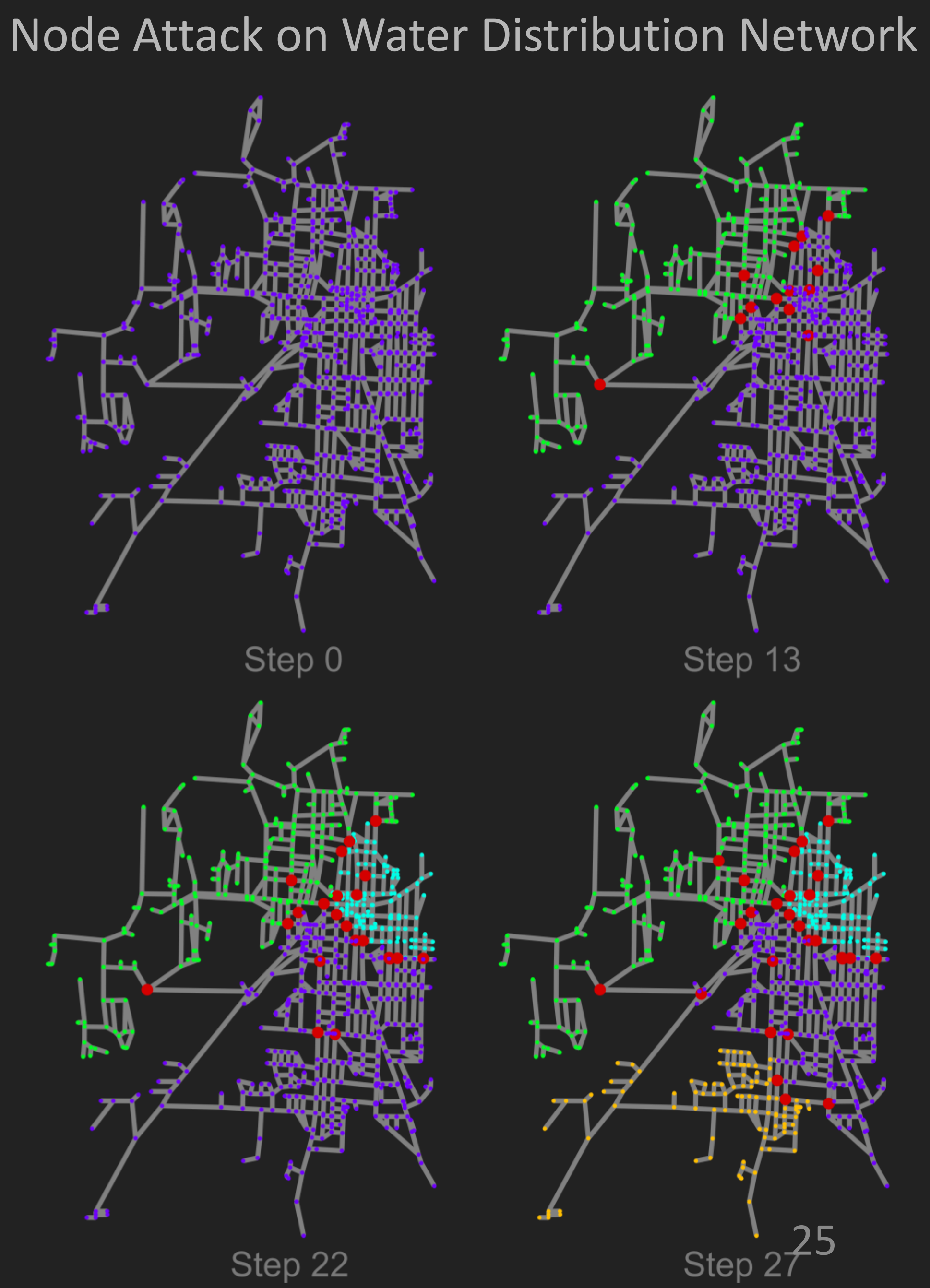
5 adjacency matrix measures +
1 approximate version

3 Laplacian matrix measures +
2 approximate versions

# 17 Failure and Attack Mechanisms

Networks can suffer from natural failures and targeted attacks.

TIGER simulates a node attack (red) on the Kentucky KY-2 water distribution network (right)



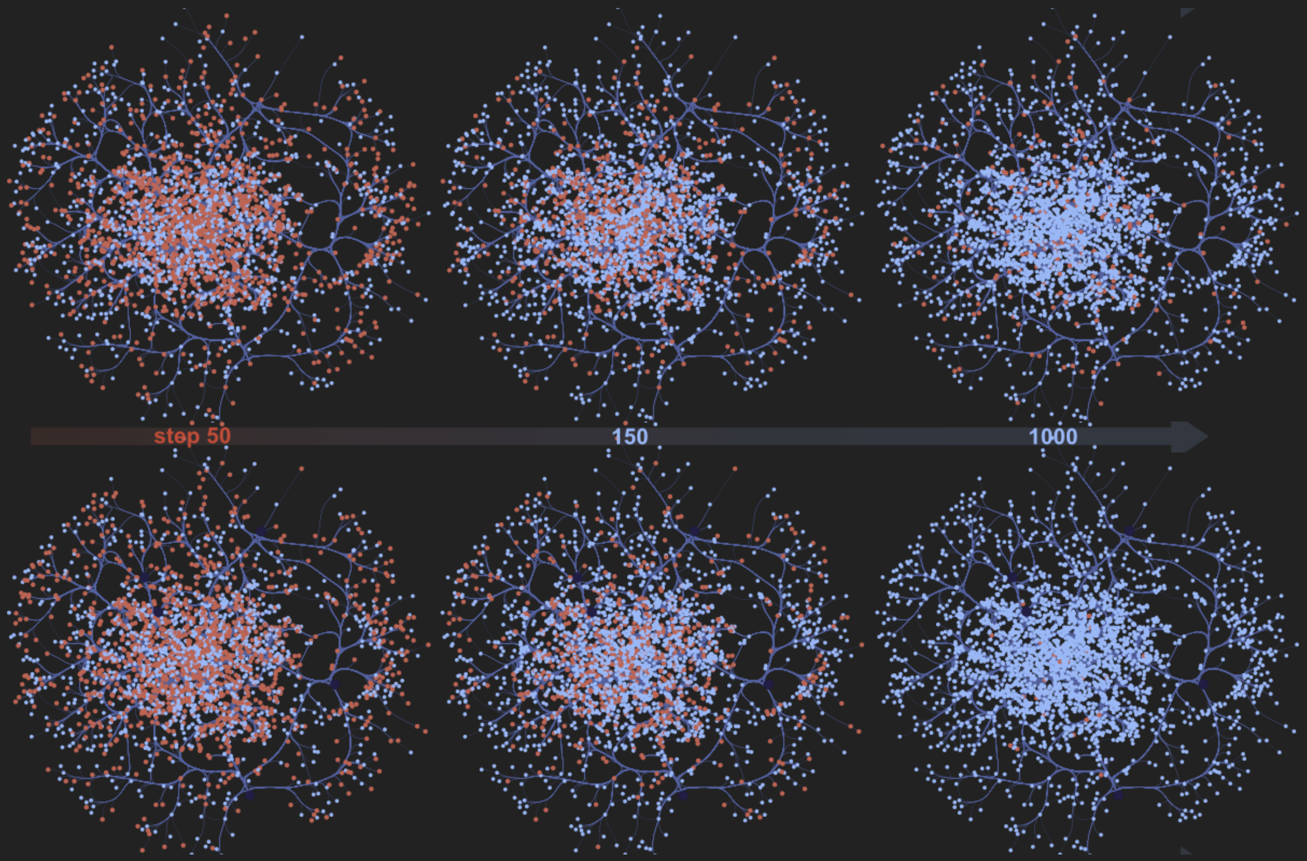
# 15 Defense Techniques

TIGER virus simulation using SIS infection model

- **Top**: no defense results in an endemic virus
- **Bottom**: defending 5 nodes with Netshield eradicates virus

Oregon-1 Autonomous System Network

**No defense**

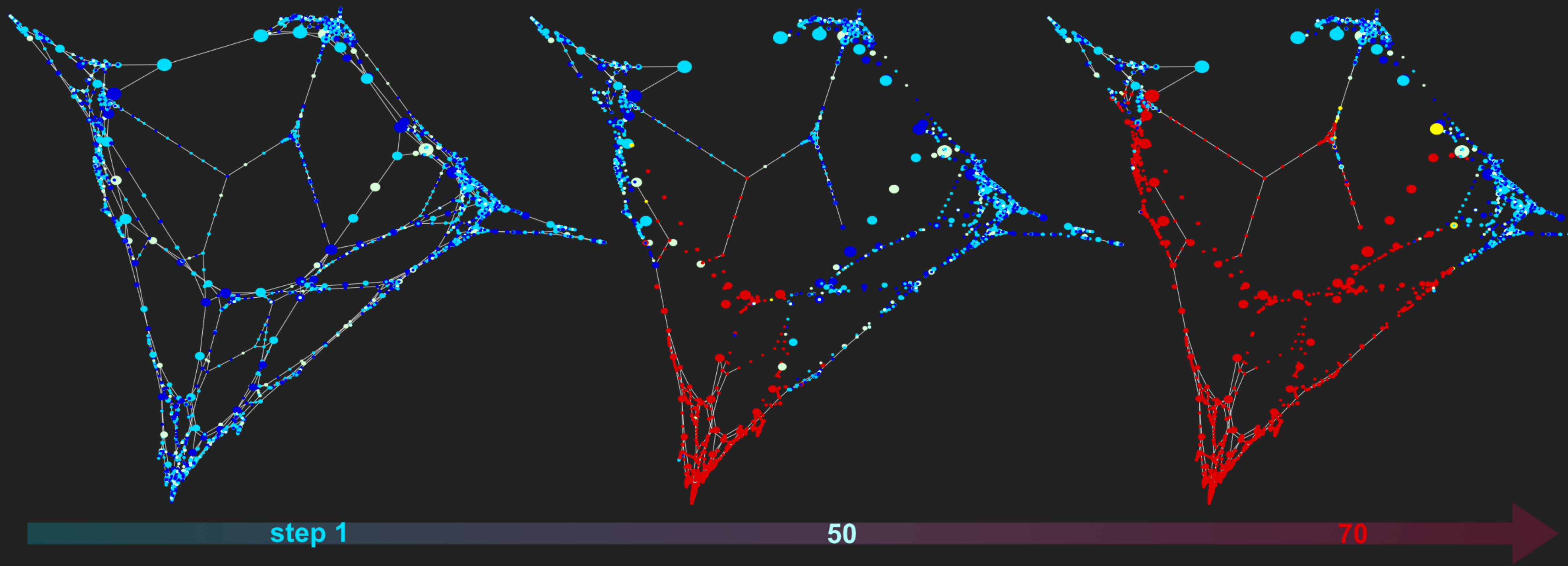



**With defense**

# Simulation Techniques

(1) entity dissemination, (2) cascading failures, (3) attacks, and (4) defenses

Example: cascading failure simulation on U.S. power grid when 4 substations are attacked



# Dissertation Research Mission

Address large-scale societal problems in cybersecurity and healthcare through **the lens of robust machine learning**

Part I:
**Tools**

**Robustness Survey** Summarize robustness literature TKDE 2021 (under review)

**TIGER** Vulnerability and robustness toolbox CIKM 2021

Part II:
**Algorithms**

**D²M** Quantify network robustness + mitigate attacks SDM 2020

Part III:
**Databases**

**MalNet-Graph** Largest cybersecurity graph database NeurIPS 2021

**MalNet-Image** Largest cybersecurity image database Submitting to CIKM 2022

Part IV:
**Models**

**UnMask** Identify robust features in images IEEE Big Data 2020

**REST** Identify robust signals in health data Web Conference 2020

# Part II: **Algorithms**

Through our survey and development of TIGER, we find that network robustness has yet to address important issues in cybersecurity

This observation motivated us to study the robustness of authentication graphs in enterprise networks

**D$^2$M** Quantify network robustness + mitigate attacks SDM 2020

# D$^2$M

# Dynamic Defense and Modeling of Adversarial Movement in Networks

SDM 2020

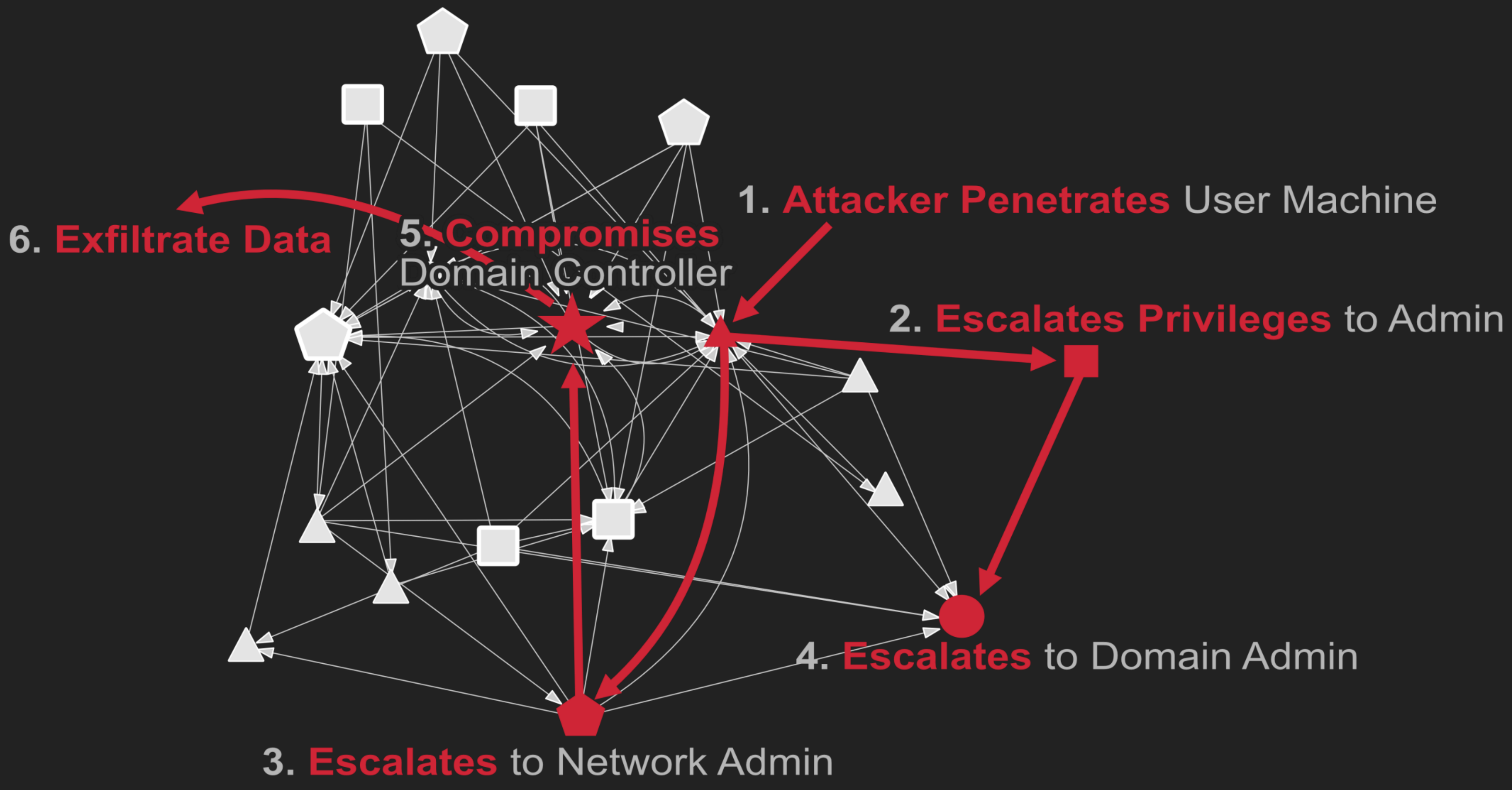

Scott Freitas
Georgia Tech

Andrew Wicker
Microsoft

Joshua Neil
Securonix

Polo Chau
Georgia Tech

# Lateral Movement Attack Chain



# Defender's Dilemma

**Goal:** Develop defense strategies and vulnerability analysis for lateral attacks

**Problem:** Sparse observable data on lateral attacks

- Ground-truth partially uncovered through investigation
- Incident reports are withheld for security and privacy
- Can not store network telemetry for more than 6 months (GDPR)
- Attackers can operate as legitimate users

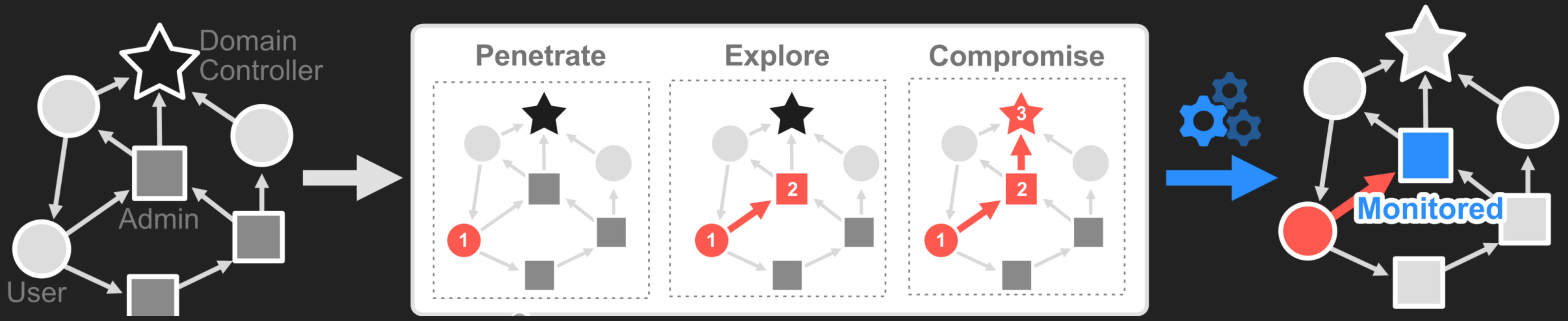
# Defender's Solution

**Goal**: Develop defense strategies and vulnerability analysis for lateral attacks

**Idea**: Simulate lateral attacks on enterprise networks

- Develop realistic lateral attack environment
- Model multiple attack strategies
- Incorporate domain knowledge

# Defender's Solution: $D^2M$ Framework




Build authentication graph

Contribution 1

**Simulate lateral attack**

Contribution 2

**Quantify Vulnerability**

Contribution 3

**Identify at-risk machine to monitor**

# Authentication Graph and Domain Knowledge

Network activity forms a directed graph *G = (V, E)*

- *V* = set of network machines
- *E* = set of edges representing authentication activity between machines
- Collect *V, E* over a period of 30 days

**Penetrate**: Attacker can start on any machine with lowest credential

**Explore and exploit**: Move randomly or using knowledge of network topology

**Exfiltrate**: Once adversary reaches domain controller, the simulation ends

# Attack Strategies: Uniformed Exploration

**Strategy 1:** RandomWalk-Explore (RWE)

- 85% chance attacker uniformly selects a neighbor
- 15% chance attacker randomly selects a $c_1$ machine; model randomness
- After visiting, attacker gains the machine's credential

# Quantifying Network Vulnerability

**Vulnerability**: risk of domain controller being compromised by lateral attack

Define as function of 3 components:

1. Network topology
2. Distribution of credentials
3. Attacker penetration point

In practice:

- Don’t know true credential distribution
- Don’t know penetration point

# Monte-Carlo To The Rescue

- Larger score = more vulnerable network
- $f(\cdot)$ simulates network attack (1=success, 0=failure)
- Sum over different penetration points
- Sum over different credential distributions
- Sum over different hygiene levels

$$L(G, h) = \frac{1}{|D_h|} \frac{1}{|R|} \sum_{\boldsymbol{d} \in \boldsymbol{D}_h} \sum_{v \in R} f(G, \boldsymbol{d}, v) \qquad L(G) = \sum_{h_i \in H} p(h_i) \cdot L(G, h_i)$$

# Identifying At-Risk Machines

Utilize network topology + attack path activity

## Strategy 1: Random Anomalous Neighbor

Vaccinate neighbors of random anomalous machines w/ weight towards recent activity

## Strategy 2: AnomalyShield

Vaccinate machines w/ high eigenvector centrality (u) and that are near anomalous activity (a)

$$AV(S_k) = \sum_{i \in S_k} \boldsymbol{u}(i) \sum_{j \in N(i)} \boldsymbol{a}(j)\boldsymbol{u}(j)$$

# Experiment Setup

Data from three networks

- 2 Microsoft tenants $G_s$, $G_l$; Los Alamos National Lab dataset $G_{lanl}$

| | $\|V\|$ | $\|E\|$ | $\rho$ | C | Avg. Degree |
|---|---|---|---|---|---|
| $G_s$ | 100 | 279 | 0.028 | 0.23 | 5.58 |
| $G_l$ | 2,039 | 3,853 | 0.001 | 0.26 | 3.78 |
| $G_{lanl}$ | 14,813 | 223,399 | 0.001 | 0.62 | 30.16 |

Network Statistics. From left to right: number of vertices $|V|$, number of edges $|E|$, density $\rho$, average clustering coefficient $C$, average out-degree of nodes in $G$.

# Quantifying Network Vulnerability

- Informed strategies lead to quicker attacks
- Improving hygiene reduces vulnerability ($h_1$=bad hygiene)
- Networks that are more well-connected are more vulnerable to lateral attacks

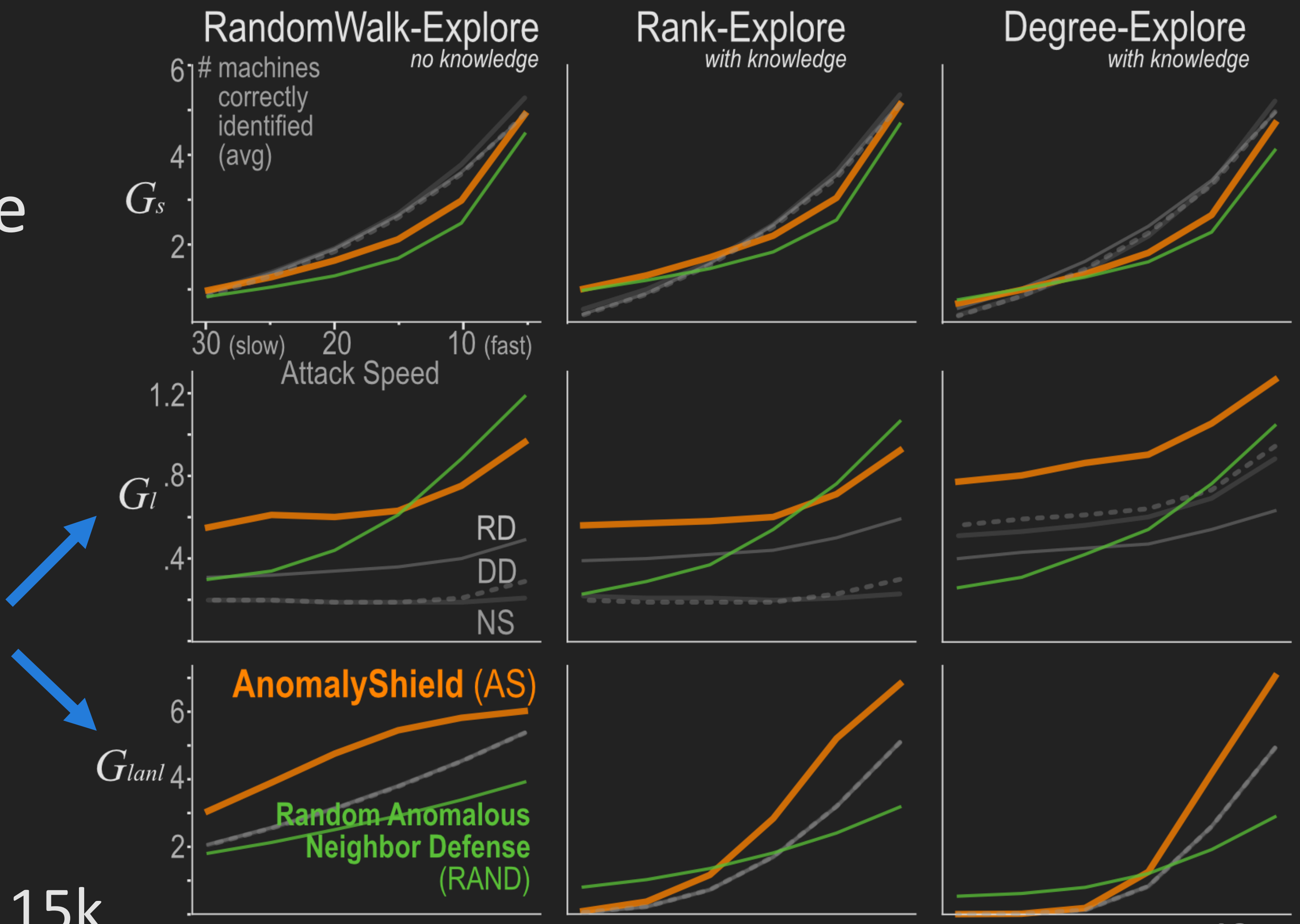
| | | Avg, Path Length | | | Vulnerability (higher = more vulnerable) | |
|---|---|---|---|---|---|---|
| Graph | Hygiene | RE | DE | RAND | L(G, h) | L(G) |
| $G_s$ | $h_1$ | 19 | 19 | 25 | .773 | .525 |
| | $h_2$ | 49 | 39 | 39 | .801 | |
| | $h_3$ | 0 | 0 | 0 | 0 | |
| $G_l$ | $h_1$ | 33 | 36 | 46 | .005 | .005 |
| | $h_2$ | 63 | 63 | 68 | .006 | |
| | $h_3$ | 133 | 139 | 139 | .004 | |
| $G_{lanl}$ | $h_1$ | 22 | 18 | 45 | .967 | .976 |
| | $h_2$ | 88 | 128 | 90 | .981 | |
| | $h_3$ | - | - | 249 | .981 | |

# Identifying At-Risk Machines

AnomalyShield an effective general defense

Larger graphs require informed defense ($G_l$, $G_{lanl}$)

$G_s$ = 100 nodes, $G_l$ = 2k, $G_{lanl}$ = 15k



# Dissertation Research Mission

Address large-scale societal problems in cybersecurity and healthcare through **the lens of robust machine learning**

Part I:
**Tools**

**Robustness Survey** Summarize robustness literature TKDE 2021 (under review)

**TIGER** Vulnerability and robustness toolbox CIKM 2021

Part II:
**Algorithms**

**D²M** Quantify network robustness + mitigate attacks SDM 2020

Part III:
**Databases**

**MalNet-Graph** Largest cybersecurity graph database NeurIPS 2021

**MalNet-Image** Largest cybersecurity image database Submitting to CIKM 2022

Part IV:
**Models**

**UnMask** Identify robust features in images IEEE Big Data 2020

**REST** Identify robust signals in health data Web Conference 2020

# Part III: **Databases**

In Part II, we focused on post-breach adversarial modeling and mitigation, in Part III our goal is to prevent lateral attacks altogether.

However, current databases are either too small or not publicly available. By creating new large-scale databases, we enable the development of next-generation malware detection models

# MalNet-Graph

## A Large-Scale Database for Graph Representation Learning

NeurIPS Datasets and Benchmarks 2021

Scott Freitas
Georgia Tech

Yuxiao Dong
Facebook AI

Joshua Neil
Securonix

Polo Chau
Georgia Tech

# MalNet Overview

Available at github.com/safreita1/malnet-graph

Part of **Nvidia Data Science Teaching Kit**

**MalNet** is a graph representation learning database with 1.2M graphs, 696 classes, and 15k nodes & 35k edges per graph

1. **Highlight** the importance of scalable graph representation learning techniques
2. **Reveal** the challenges of working with highly imbalanced graph data
3. **Showcase** the effectiveness of simple baselines on non-attributed graphs
4. **Enable** new research into imbalanced classification, explainability, and the impact of class hardness

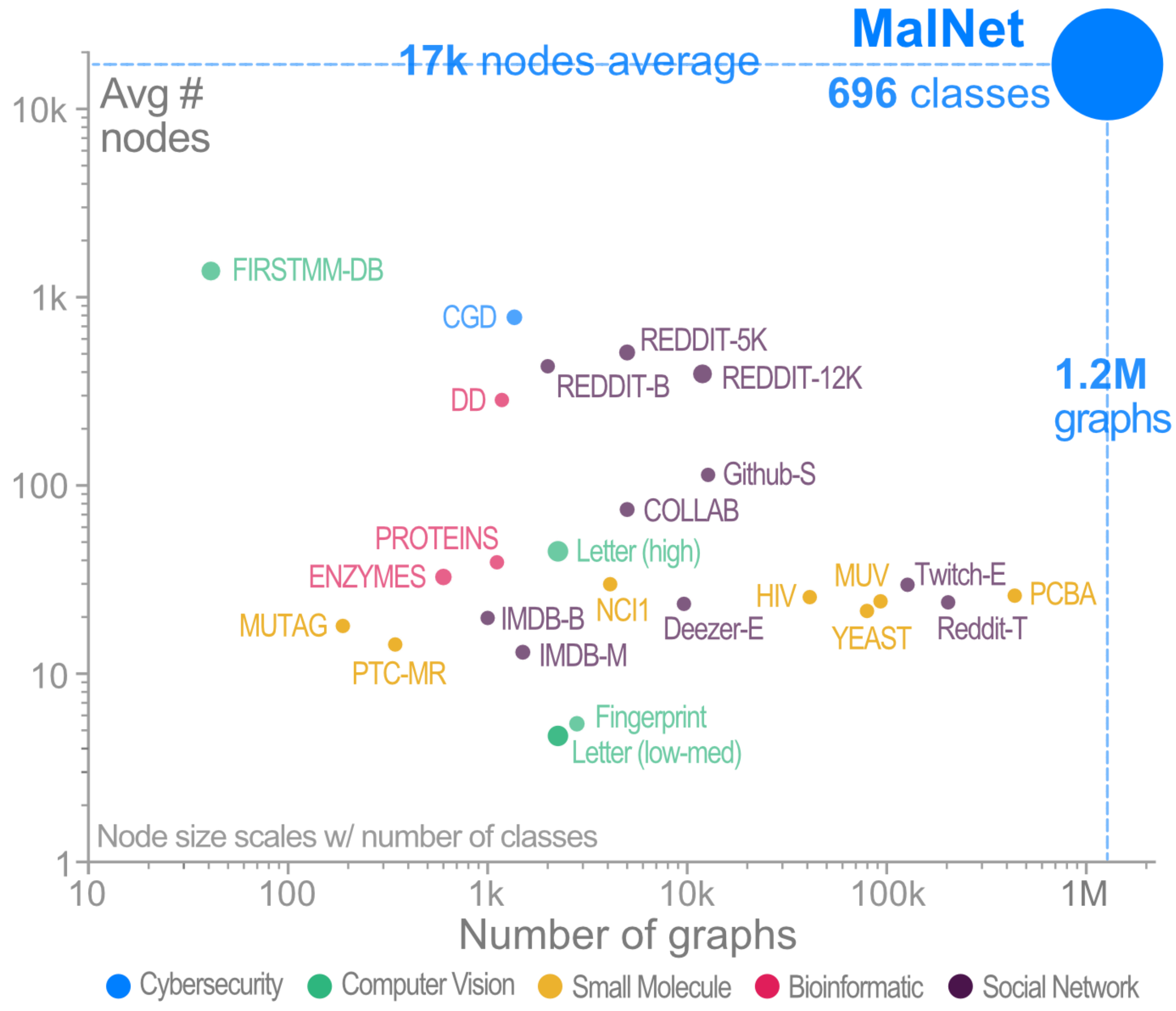
# Why MalNet-Graph?

Number of limitations with existing graph classification databases

1. contain relatively few graphs
2. small graphs, terms nodes and edges
3. limited number of classes

Compared to the popular REDDIT-12K database, we offer

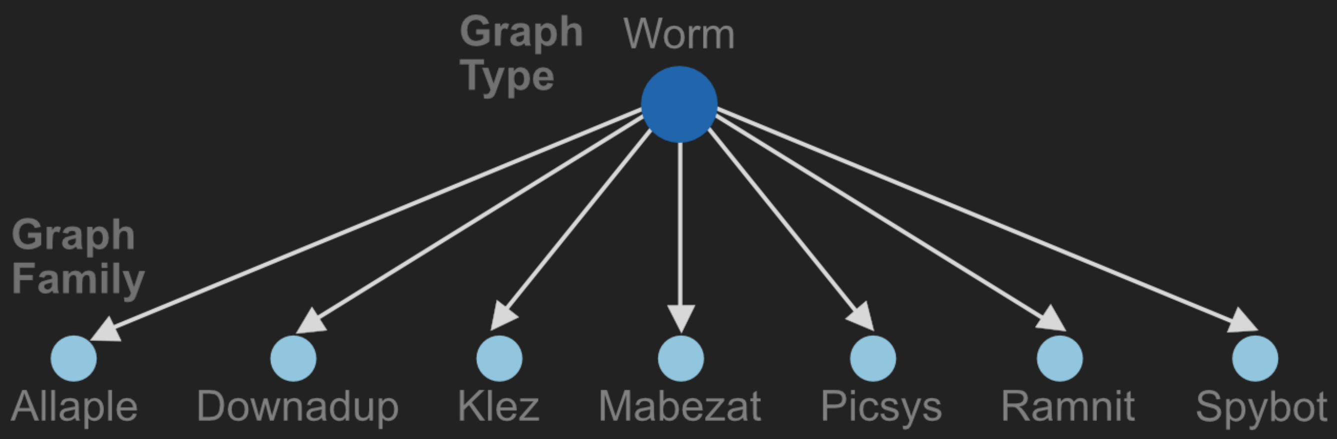
1. 105x more graphs
2. 44x larger graphs on average
3. 63x more classes



# Collecting Candidate Graphs

Select Android ecosystem for

1. large market share
2. easy accessibility
3. diversity of malware




Collecting took 1 month and 10TB of storage

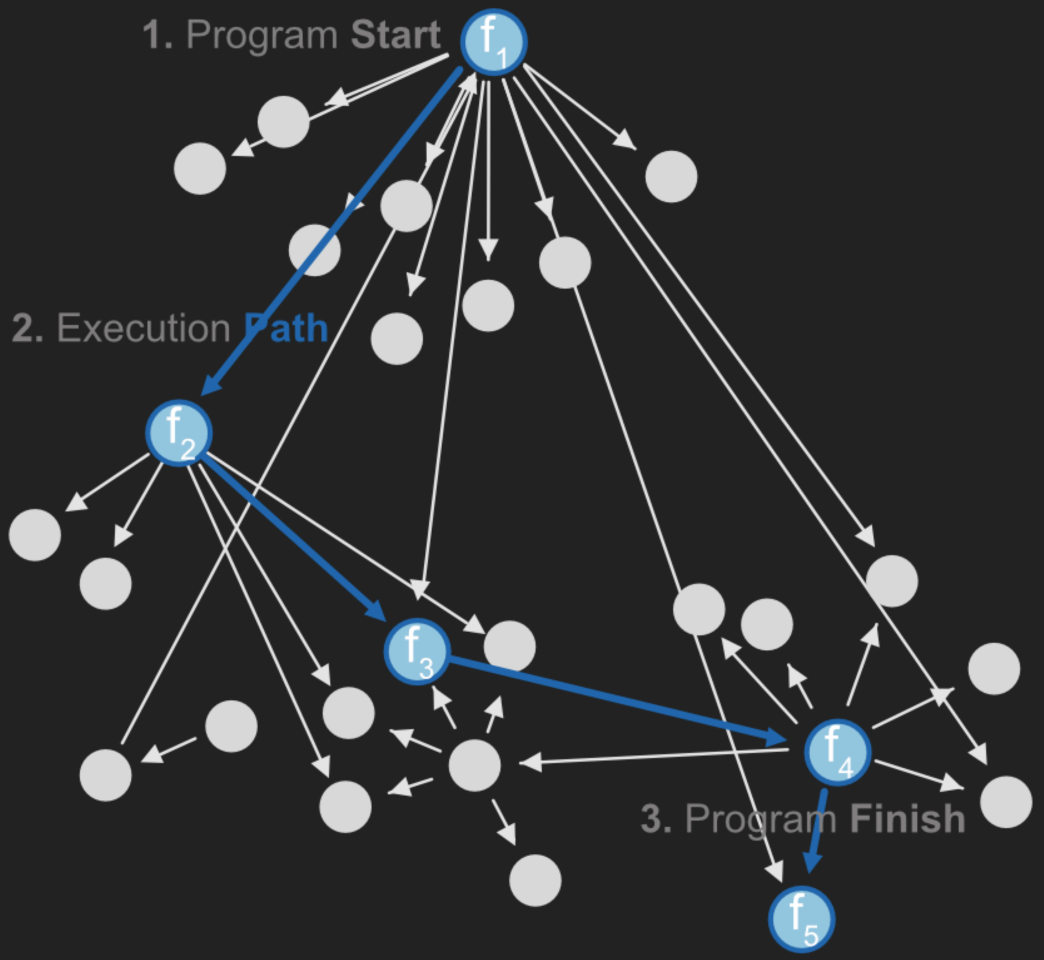
1. Collect Android APK files from AndroidZoo
2. Select files with *type* and *family* labels
3. Collect raw VirusTotal reports for each file

# Processing the Graphs

Extract function call graph (FCG) from APK

Directed graph containing disconnected components and isolates

443GB of disk space; edge list format for wide support, readability, and ease of use




Example function call graph

# MalNet for New Research and Discoveries

## Revealing new discoveries

1. Graph representation learning scalability and large class imbalance issues
2. Simple baselines are surprisingly effective
3. GNNs are not state-of-the-art

## Enabling new research directions

1. Imbalanced classification
2. Explainability
3. Class hardness

# Experiment Setup

- Split MalNet-Graph data 70/10/20 for training, validation, test
- Create MalNet Tiny containing 5k graphs across 5 balanced classes
- Evaluate 7 state-of-the-art methods using macro-F1 score
    - 2 GNNs (GCN, GIN ) and 5 DM techniques (LDP, NoG, Feather, Slaq-VNGE, Slaq-LSD)
- Each GNN has a parameter search over lr and hidden units
    - Took 26 days to complete on Nvidia DGX A100
- DM techniques have parameter search over method and RF model

# Graph Classification Experiments

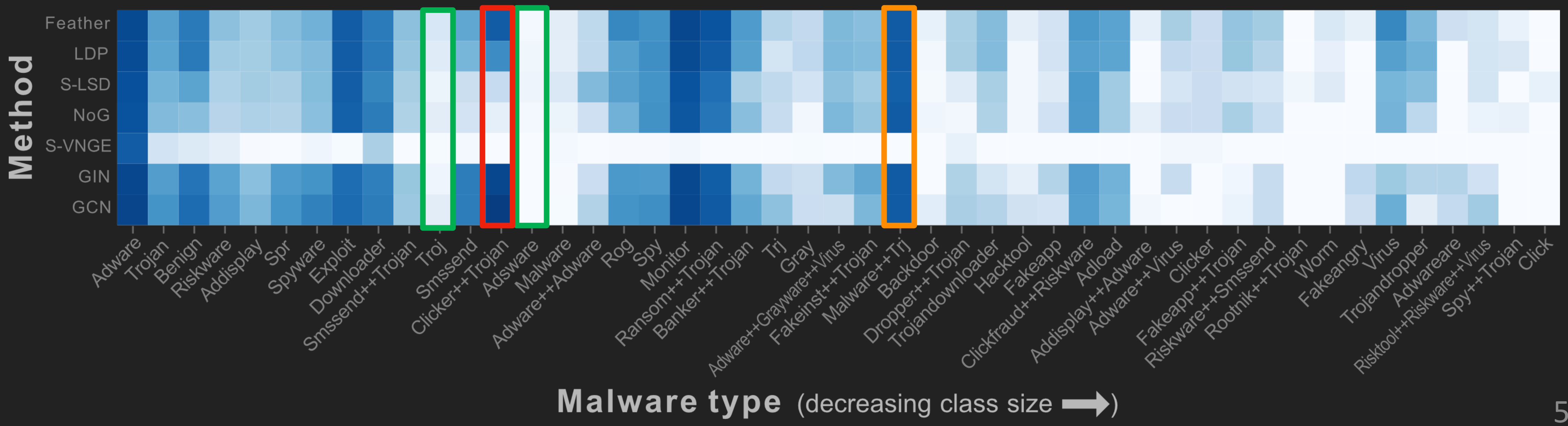
- Less diversity, better performance
- Simple baselines surprisingly effective
- GNNs not state-of-the-art

| Method | Type (47 classes) | | | Family (696 classes) | | | Tiny |
|---|---|---|---|---|---|---|---|
| | Macro-F1 | Precision | Recall | Macro-F1 | Precision | Recall | Accuracy |
| Feather | .41 | .71 | .35 | .34 | .56 | .29 | .86 |
| LDP | .38 | .69 | .31 | .34 | .55 | .28 | .86 |
| GIN | .39 | .57 | .36 | .28 | .32 | .28 | **.90** |
| GCN | .38 | .51 | .35 | .21 | .24 | .21 | .81 |
| Slaq-LSD | .33 | .62 | .26 | .24 | .42 | .19 | .76 |
| NoG | .30 | .62 | .25 | .25 | .42 | .21 | .77 |
| Slaq-VNGE | .04 | .07 | .04 | .01 | .01 | .01 | .53 |

# Graph Classification Experiments

**Class hardness exploration**—*Malware++Trj* outperforms *Troj* and *Adsware*, which contain many more examples

**Explainability research**—why does Feather, GIN, GCN outperform on Clicker++Trojan



# MalNet-Image

## A Large-Scale Image Database of Malicious Software

Submitting to CIKM 2022


Scott Freitas
Georgia Tech

Rahul Duggal
Georgia Tech

Polo Chau
Georgia Tech

# Why MalNet-Image?

- Over 1.2M images across 47 types and 696 families
- **24x** more images, **70x** more classes compared to the next largest database
- Enable large-scale malware detection and classification research

| | Dataset | Images | Classes |
|---|---|---|---|
| Public | **MalNet** | **1,262,024** | **696** |
| | Virus-MNIST | 51,880 | 10 |
| | Malimg | 9,458 | 25 |
| Private | Stamina | 782,224 | 2 |
| | McAfee | 367,183 | 2 |
| | Kancherla | 27,000 | 2 |
| | Choi | 12,000 | 2 |
| | Fu | 7,087 | 15 |
| | Han | 1,000 | 50 |
| | IoT DDoS | 365 | 3 |

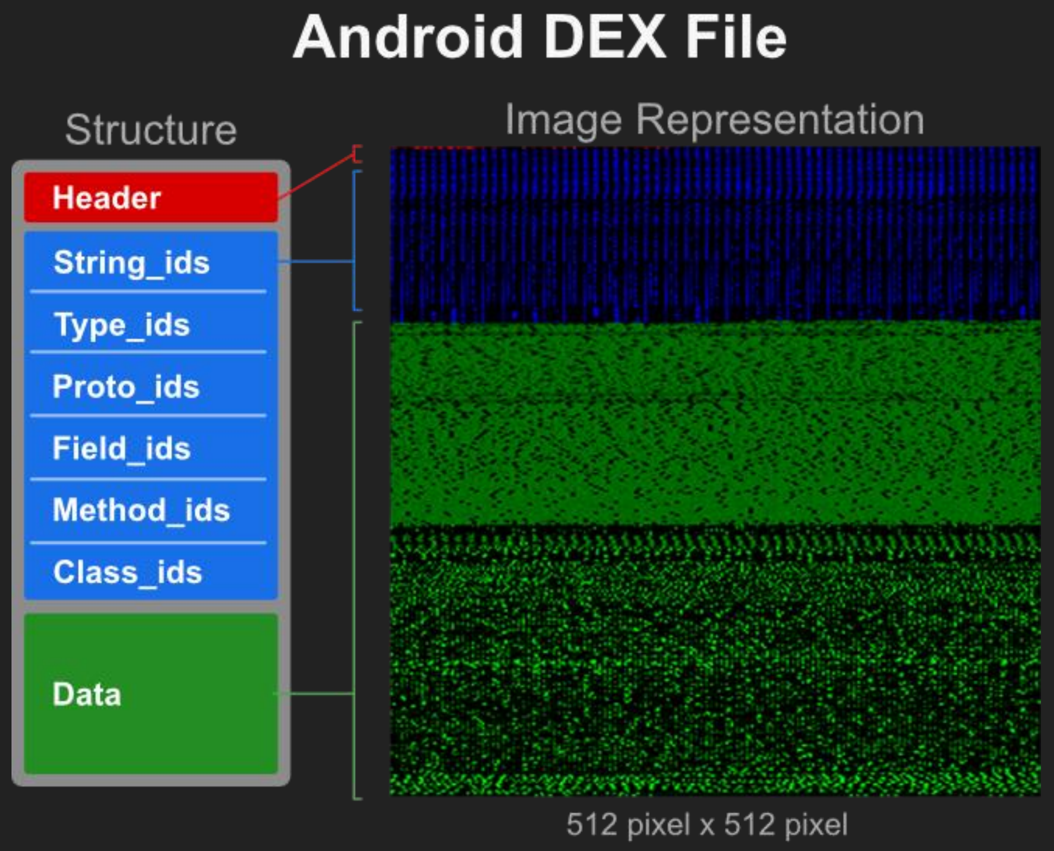
# MalNet-Image Construction

## Collecting candidate images

- APK files from AndroZoo repository
- 1-1 match with FCGs from MalNet-Graph

## Processing the images

- Extract DEX file from each APK
- Convert DEX to 1D array of 8-bit integers
- Convert 1D array to 2D array
- RGB color coded channels represent structure information



# Experiment Setup

- MalNet-Image data split 70/10/20 for training, validation, test
  - Stratified split across binary, type and family labels, respectively
- Evaluate 7 DL models using macro-F1 score
  - 3 ResNet (18, 50, 101), 2 DenseNet (121, 169), and 2 MobileNetV2 (x.5, x1)
- Each model is trained for 100 epochs using an Adam optimizer
- Experiments run on an Nvidia DGX-1 containing 8 V100 GPUs

# Benchmarking Techniques

We evaluate numerous malware detection and classification techniques, previously studied on private or small-scale databases, such as

- Semantic information encoding via colored channels vs grayscale
- Deep learning model architectures (ResNet, DenseNet, MobileNetV2)
- Model pretraining on ImageNet versus training from scratch
- Imbalanced classification techniques (focal loss, class reweighting)

Select ResNet-18 model trained from scratch on grayscale images using CE loss and class reweighting due to strong performance and quick training

# Malware Classification Capabilities

- Promising malware detection results
- Type level performance on-par with family level
- Image models significantly outperform FCG based methods

| Model | Params | MFlops | Binary | | | Type (47 classes) | | | Family (696 classes) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | F1 | Prec. | Recall | M-F1 | Prec. | Recall | M-F1 | Prec. | Recall |
| **ResNet18** | **12M** | **1,820** | **.86** | **.89** | **.84** | **.47** | **.56** | **.42** | **.45** | **.54** | **.42** |
| ResNet50 | 26M | 3,877 | .85 | .91 | .81 | .48 | .57 | .44 | .47 | .54 | .44 |
| ResNet101 | 45M | 7,597 | .86 | .88 | .84 | .48 | .59 | .44 | .47 | .54 | .44 |
| DenseNet121 | 7.9M | 2,872 | .86 | .90 | .83 | .47 | .56 | .43 | .46 | .53 | .44 |
| DenseNet169 | 14M | 3,403 | .86 | .89 | .84 | .48 | .57 | .43 | .46 | .53 | .43 |
| MobileNetV2 (x.5) | 1.9M | 100 | .86 | .89 | .83 | .46 | .55 | .42 | .45 | .53 | .42 |
| MobileNetV2 (x1) | 3.5M | 329 | .85 | .89 | .83 | .45 | .53 | .42 | .44 | .53 | .41 |

# Dissertation Research Mission

Address large-scale societal problems in cybersecurity and healthcare through **the lens of robust machine learning**

Part I:
**Tools**

**Robustness Survey** Summarize robustness literature TKDE 2021 (under review)

**TIGER** Vulnerability and robustness toolbox CIKM 2021

Part II:
**Algorithms**

**D$^2$M** Quantify network robustness + mitigate attacks SDM 2020

Part III:
**Databases**

**MalNet-Graph** Largest cybersecurity graph database NeurIPS 2021

**MalNet-Image** Largest cybersecurity image database Submitting to CIKM 2022

Part IV:
**Models**

**UnMask** Identify robust features in images IEEE Big Data 2020

**REST** Identify robust signals in health data Web Conference 2020

# Part IV: Models

Having access to large-scale robust data, doesn’t guarantee model robustness. Therefore, we focus on developing robust models

Specifically, our goal is to tackle two high-impact societal problems in **cybersecurity** and **healthcare** affecting millions of lives—*through the lens of robust deep learning models*

# UnMask

## Adversarial Detection and Defense Through Robust Feature Alignment

IEEE Big Data 2020

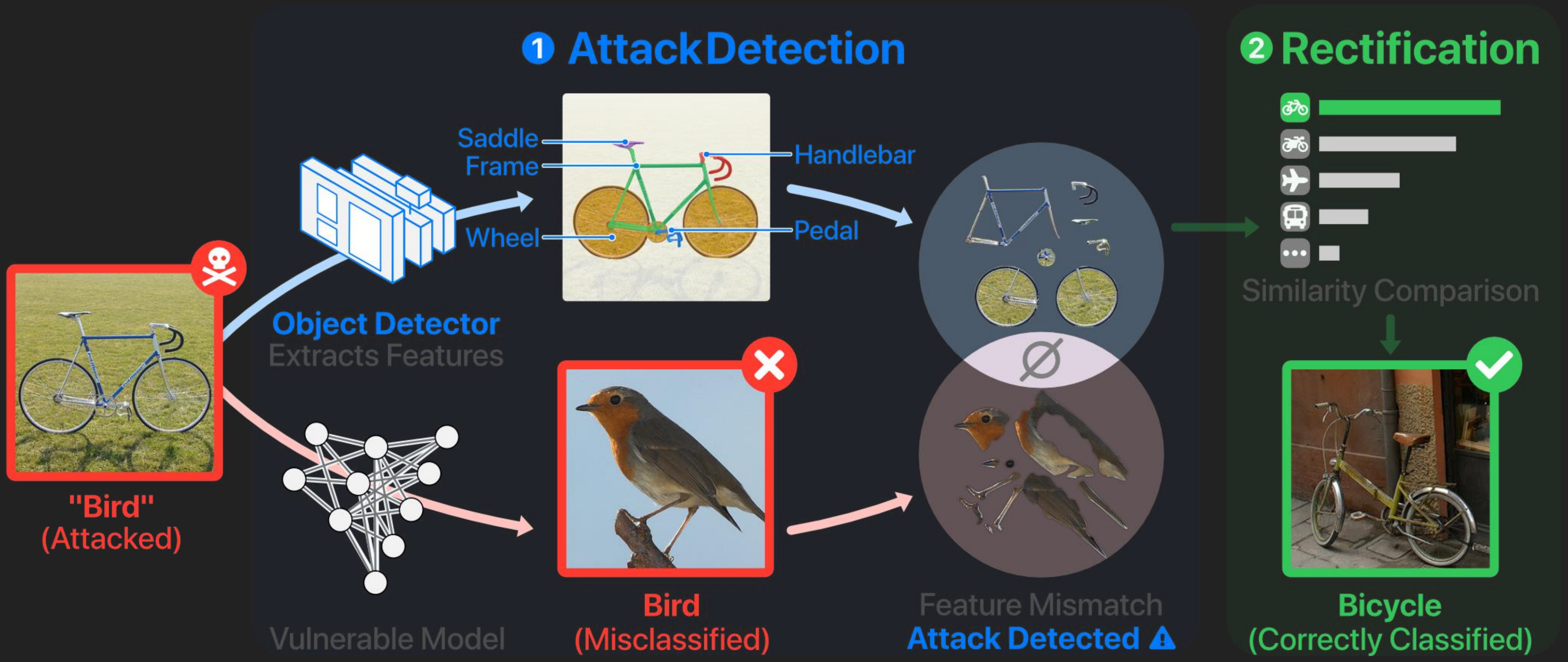

Scott Freitas
Georgia Tech

Shang-Tse Chen
National Taiwan University

Polo Chau
Georgia Tech

Jay Wang
Georgia Tech

# Protection Via Robust Feature Alignment



# Robust Features → Robust Model

Robustness is a function of the data [Ilyas 2019]

- Adversarial examples attributed to non-robust features
- Non-robust features predictive but not human comprehensible
- Training **only** on robust features significantly lowers benign accuracy

# Image’s Robust Features


wheel,
frame,
seat,
…

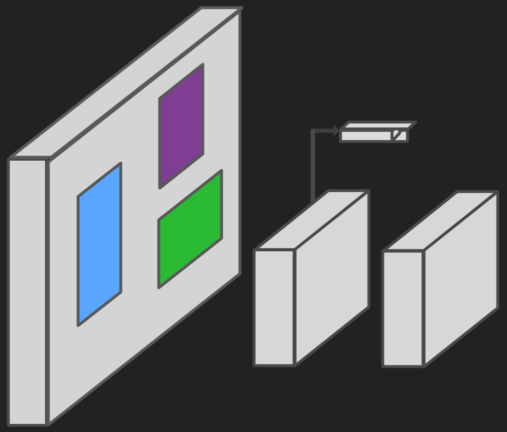
Bike’s Robust Features

- Human comprehensible
- Forms foundation for adversarial defense

# Extracting the Robust Features

Obtaining robust features

- Mask-RCNN trained on segmentation masks
- One mask per robust feature
- Dataset has 44 robust features
- e.g., **Bike**: wheel, seat, frame



Mask R-CNN

# Experiment Setup: Datasets

## Building-block extractor (Mask R-CNN)

- Based on Feature Pyramid Network and ResNet101
- Trained and evaluated using PASCAL-Part dataset

## Vulnerable CNN models

- ResNet50, DenseNet121
- Trained on PASCAL-VOC 2010; evaluated on Flickr

## Detection and defense

- Flickr, matching PASCAL-VOC classes

# Experiment Setup: Evaluation

Four adversarial attacks

- PGD-$L_\infty$, PGD-$L_2$, MI-FGSM $L_\infty$, MI-FGSM $L_2$

Four levels of feature overlap

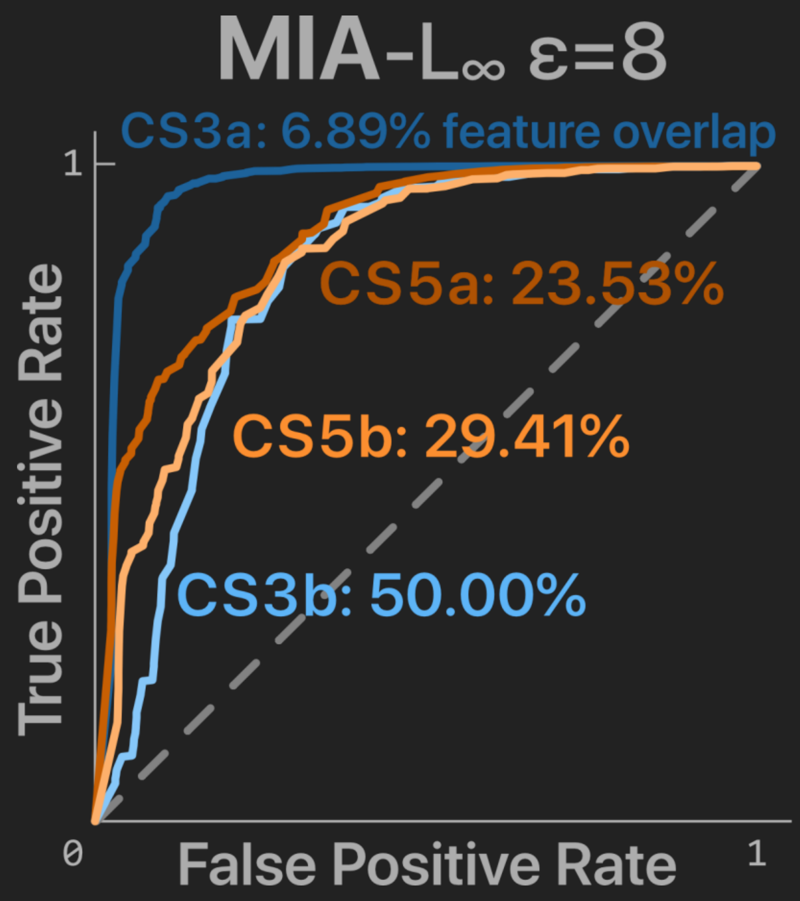
- Test's effectiveness of robust feature alignment in different setups

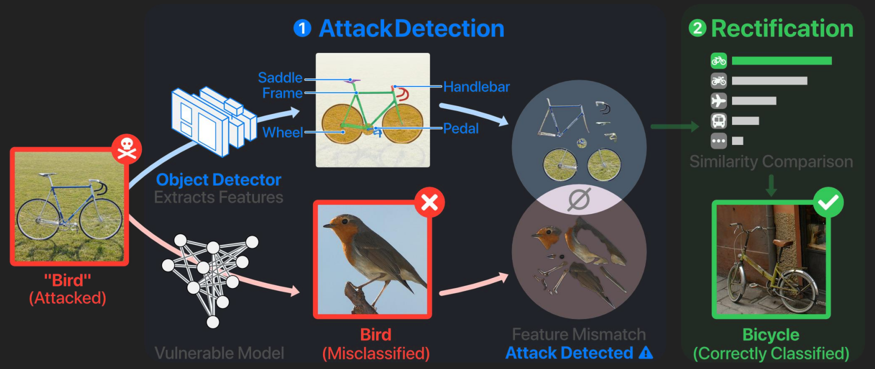
| **Class Set** | **Classes** | **Unique Parts** | **Overlap** |
|---|---|---|---|
| CS3a | 3 | 29 | 6.89% |
| CS3b | 3 | 18 | 50.00% |
| CS5a | 5 | 34 | 23.53% |
| CS5b | 5 | 34 | 29.41% |

# Detecting Attacks

Evaluating detection of adversarial images

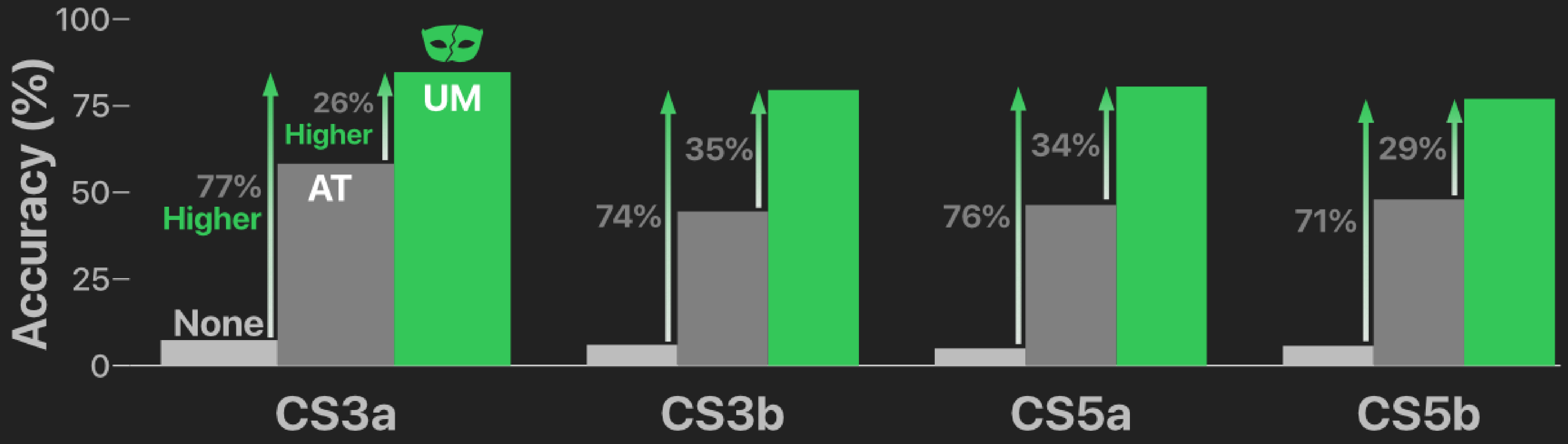
- 50:50 ratio of benign/adversarial
- Low feature overlap, better performance
- Feature selection more important than number of features



# Countering Attacks

## UnMask outperforms adversarial training



# REST

## Robust and Efficient Neural Networks for Sleep Monitoring in the Wild

Web Conference (WWW) 2020


Rahul Duggal*
Georgia Tech

Cao Xiao
Amplitude

Jimeng Sun
UIUC

Scott Freitas*
Georgia Tech

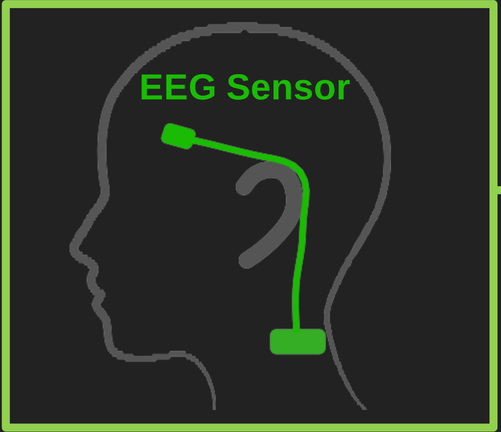
Polo Chau
Georgia Tech

*Equal contribution

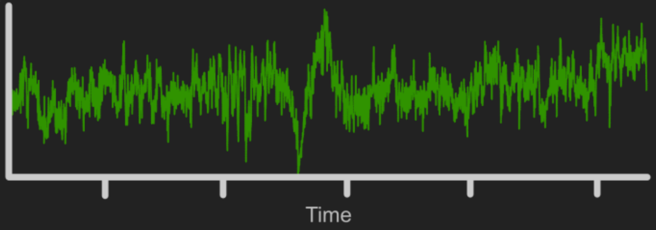
# Importance of Sleep Diagnosis

Imperative to develop **accurate** and **efficient** sleep assessments

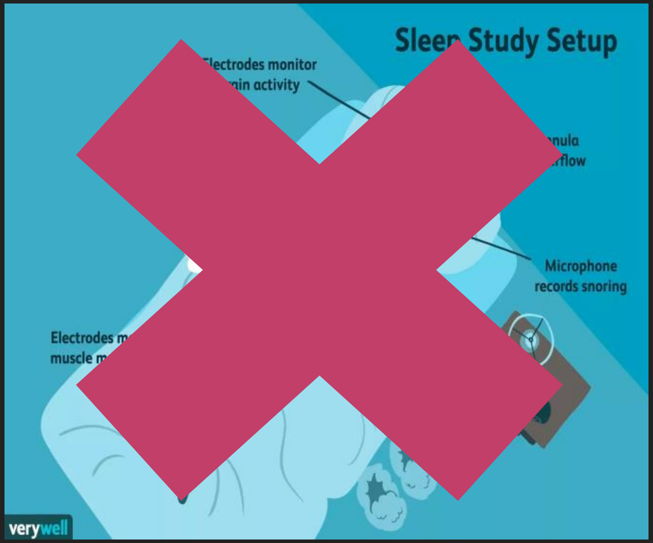
# Sleep Diagnosis Workflow



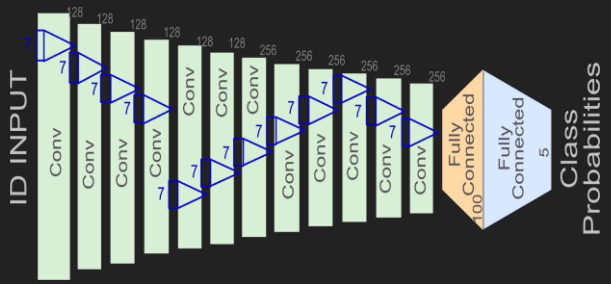

1. At home monitoring




1. Overnight sleep exam

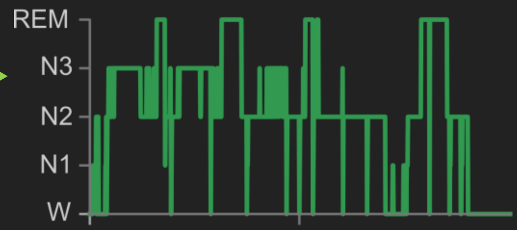



2. Deep neural network




- Costly
- Invasive
- Inconvenient

3. Hypnogram

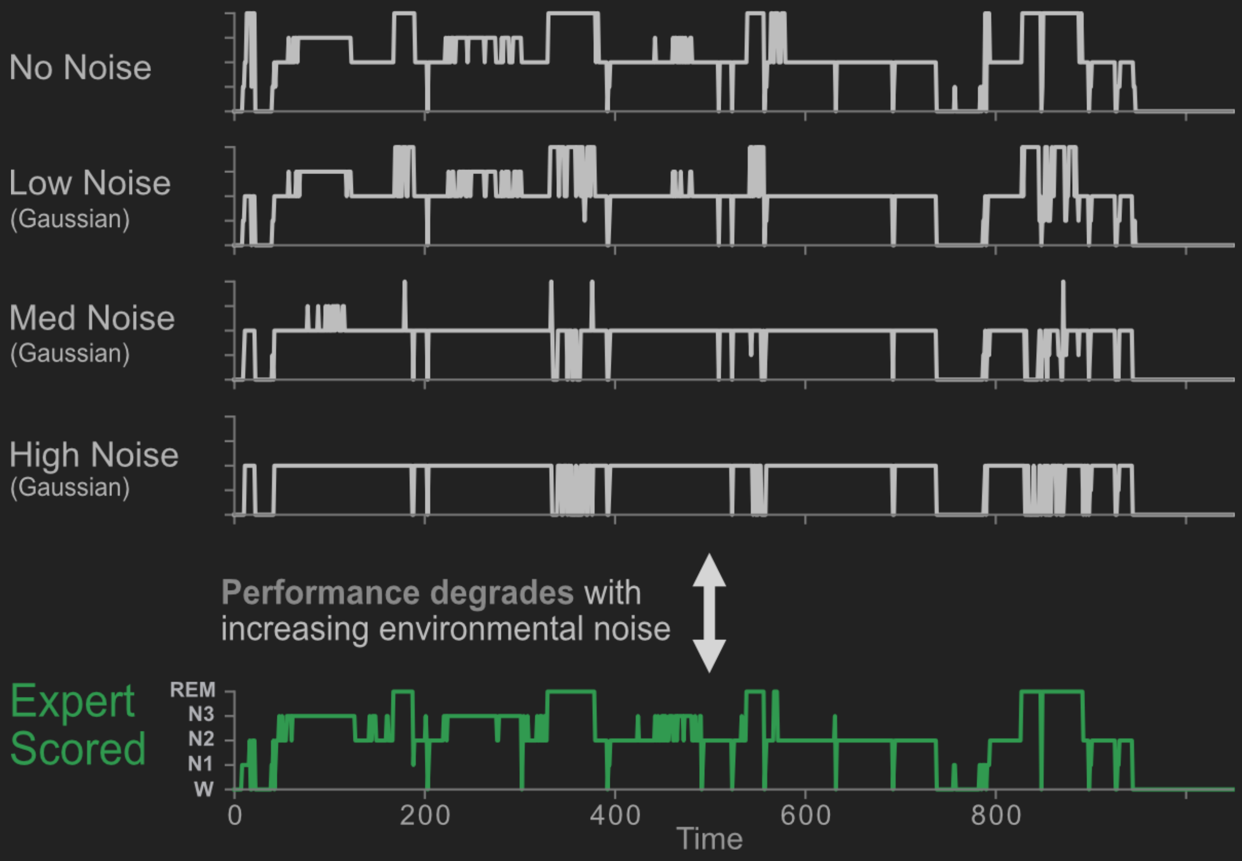



4. Team of doctors

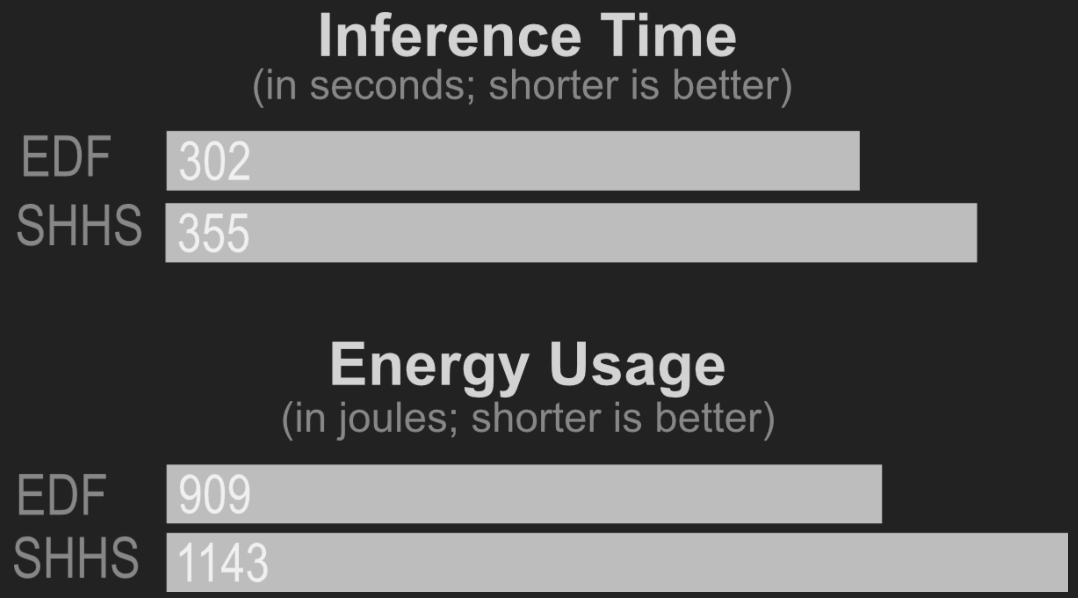
Diagnosis

# Key Challenges

## Susceptibility to noise



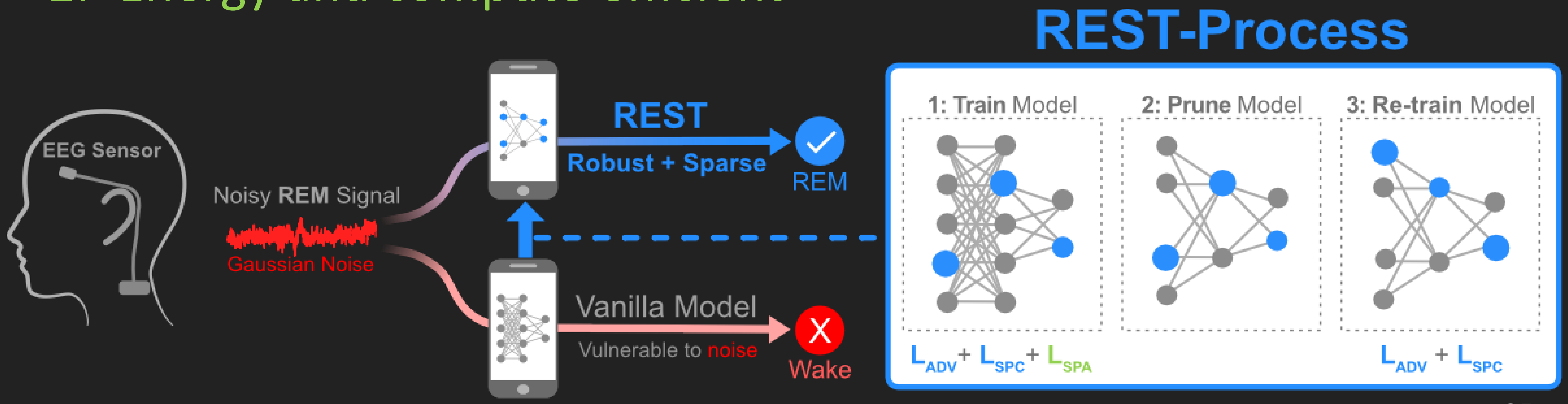

## Latency and energy use



# REST Process

Develop **neural networks** for **home sleep monitoring** that are

1. Robust to noise
2. Energy and compute efficient



# Evaluation: Setup

## Datasets

Sleep-EDF

- Collected at **home**
- More noisy

SHHS

- Collected in **sleep lab**
- Less noisy

## Metrics

Noise robustness

- Macro-F1 score avg. over test patients

Efficiency

- FLOPS to score one EEG input
- Inference time to score one night
- Joules to score one night

Measured on
Pixel-2 phone

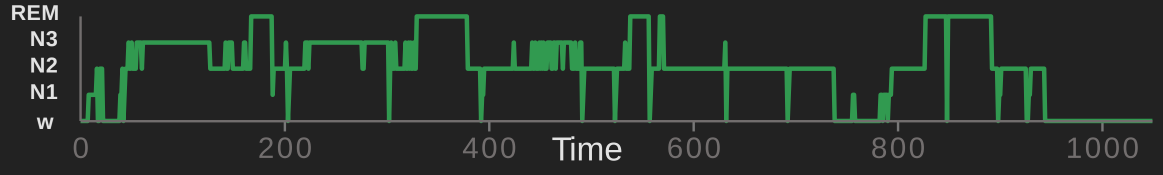
# Noise Robustness

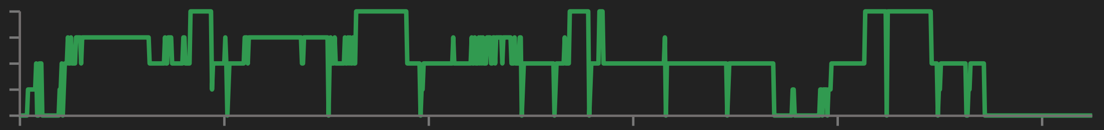
REST models perform well in noisy environments

| Data | Method | Compress | No noise | Adversarial Low | Adversarial Med | Adversarial High | Gaussian Low | Gaussian Med | Gaussian High | Shot Low | Shot Med | Shot High |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sleep-EDF | Sors [32] | ✗ | 0.67 ± 0.02 | 0.57 ± 0.02 | 0.51 ± 0.04 | 0.19 ± 0.06 | 0.66 ± 0.03 | 0.60 ± 0.03 | 0.39 ± 0.08 | 0.58 ± 0.04 | 0.42 ± 0.0[illegible] | 0.11 ± 0.03 |
| | Liu [26] | ✓ | **0.69** ± 0.02 | 0.52 ± 0.07 | 0.41 ± 0.07 | 0.09 ± 0.02 | 0.67 ± 0.02 | 0.53 ± 0.02 | 0.28 ± 0.04 | 0.52 ± 0.03 | 0.31 ± 0.04 | 0.06 ± 0.01 |
| | Blanco [7] | ✓ | 0.68 ± 0.01 | 0.51 ± 0.06 | 0.40 ± 0.06 | 0.09 ± 0.02 | 0.65 ± 0.02 | 0.54 ± 0.04 | 0.31 ± 0.10 | 0.53 ± 0.04 | 0.34 ± 0.09 | 0.08 ± 0.02 |
| | Ford [15] | ✓ | 0.64 ± 0.01 | 0.59 ± 0.01 | 0.60 ± 0.02 | 0.31 ± 0.08 | 0.65 ± 0.01 | 0.67 ± 0.02 | 0.57 ± 0.03 | 0.67 ± 0.02 | 0.60 ± 0.02 | 0.10 ± 0.01 |
| | REST (A) | ✓ | 0.66 ± 0.02 | 0.64 ± 0.02 | 0.64 ± 0.02 | 0.61 ± 0.02 | 0.66 ± 0.02 | 0.67 ± 0.01 | 0.66 ± 0.01 | 0.67 ± 0.01 | 0.66 ± 0.01 | **0.42** ± 0.06 |
| | REST (A+S) | ✓ | **0.69** ± 0.01 | **0.67** ± 0.02 | **0.66** ± 0.01 | **0.61** ± 0.03 | **0.69** ± 0.01 | **0.68** ± 0.01 | **0.67** ± 0.02 | **0.68** ± 0.01 | **0.67** ± 0.02 | **0.42** ± 0.08 |

# Noise Robustness

## Expert Scored (ground truth)

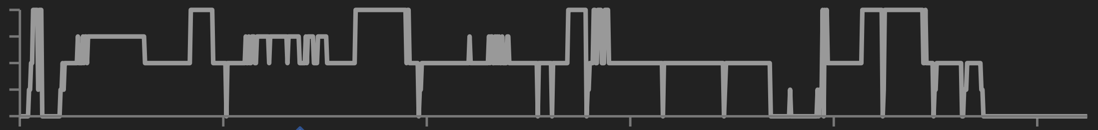


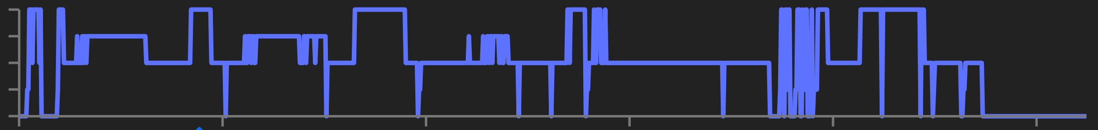



## State-of-the-Art Model

## Rest(A+S) Model

No Noise





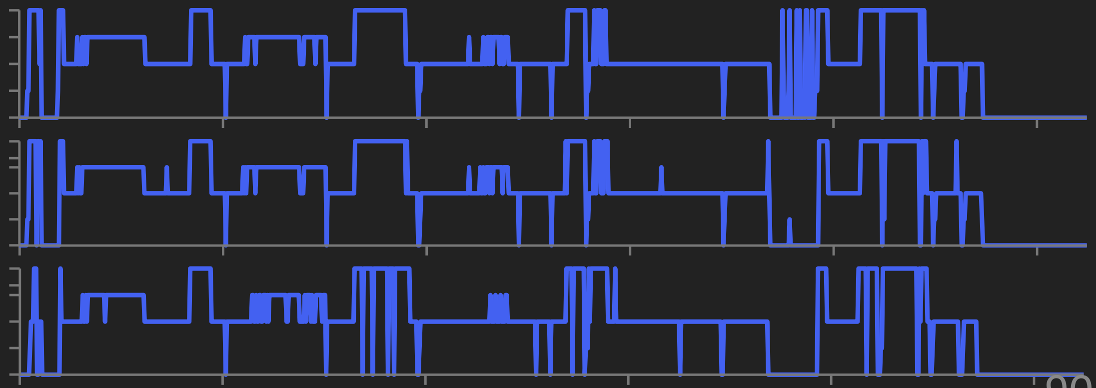
Accuracy of Hypnogram decreases with increasing noise in EEG

Accuracy of Hypnogram persists with increasing noise in EEG

Low Noise (Gaussian)

Med Noise (Gaussian)

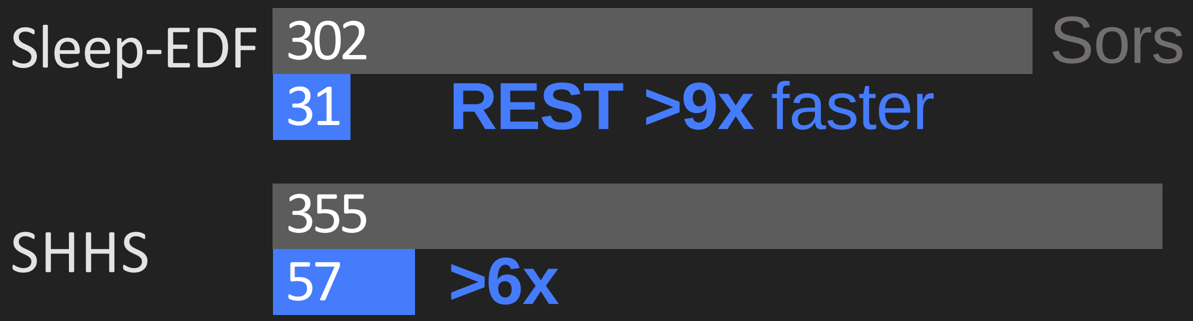
High Noise (Gaussian)



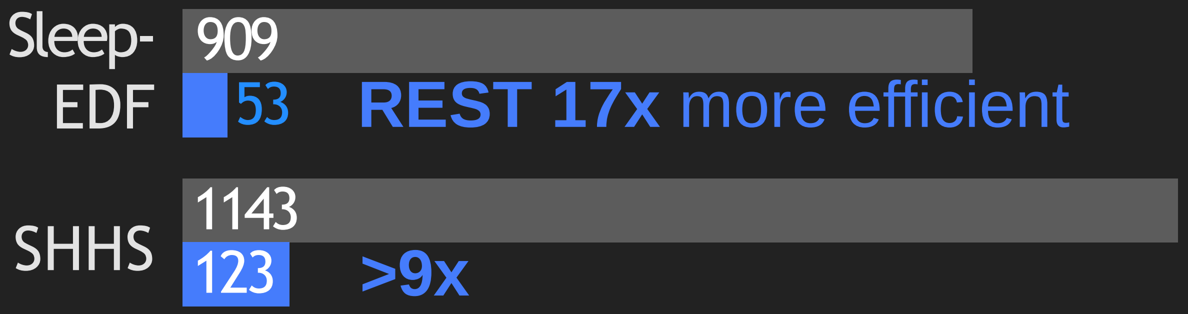
# Model Efficiency

Performance evaluated on pixel 2 smartphone

**Inference Time**
(in seconds; shorter is better)




**Energy Usage**
(in joules; shorter is better)




REST is **faster** and **efficient**

Thanks!

# Dissertation Research Mission

Address large-scale societal problems in cybersecurity and healthcare through **the lens of robust machine learning**

Part I:
**Tools**

**Robustness Survey** Summarize robustness literature TKDE 2021 (under review)
**TIGER** Vulnerability and robustness toolbox CIKM 2021

Part II:
**Algorithms**

**D$^2$M** Quantify network robustness + mitigate attacks SDM 2020

Part III:
**Databases**

**MalNet-Graph** Largest cybersecurity graph database NeurIPS 2021
**MalNet-Image** Largest cybersecurity image database Submitting to CIKM 2022

Part IV:
**Models**

**UnMask** Identify robust features in images IEEE Big Data 2020
**REST** Identify robust signals in health data Web Conference 2020