Deep Learning in Cyber Security

Dr. Manas Khatua
Associate Professor
Dept. of Computer Science & Engineering
Indian Institute of Technology Guwahati
URL: http://manaskhatua.github.io/
Email: manaskhatua@iitg.ac.in

Content

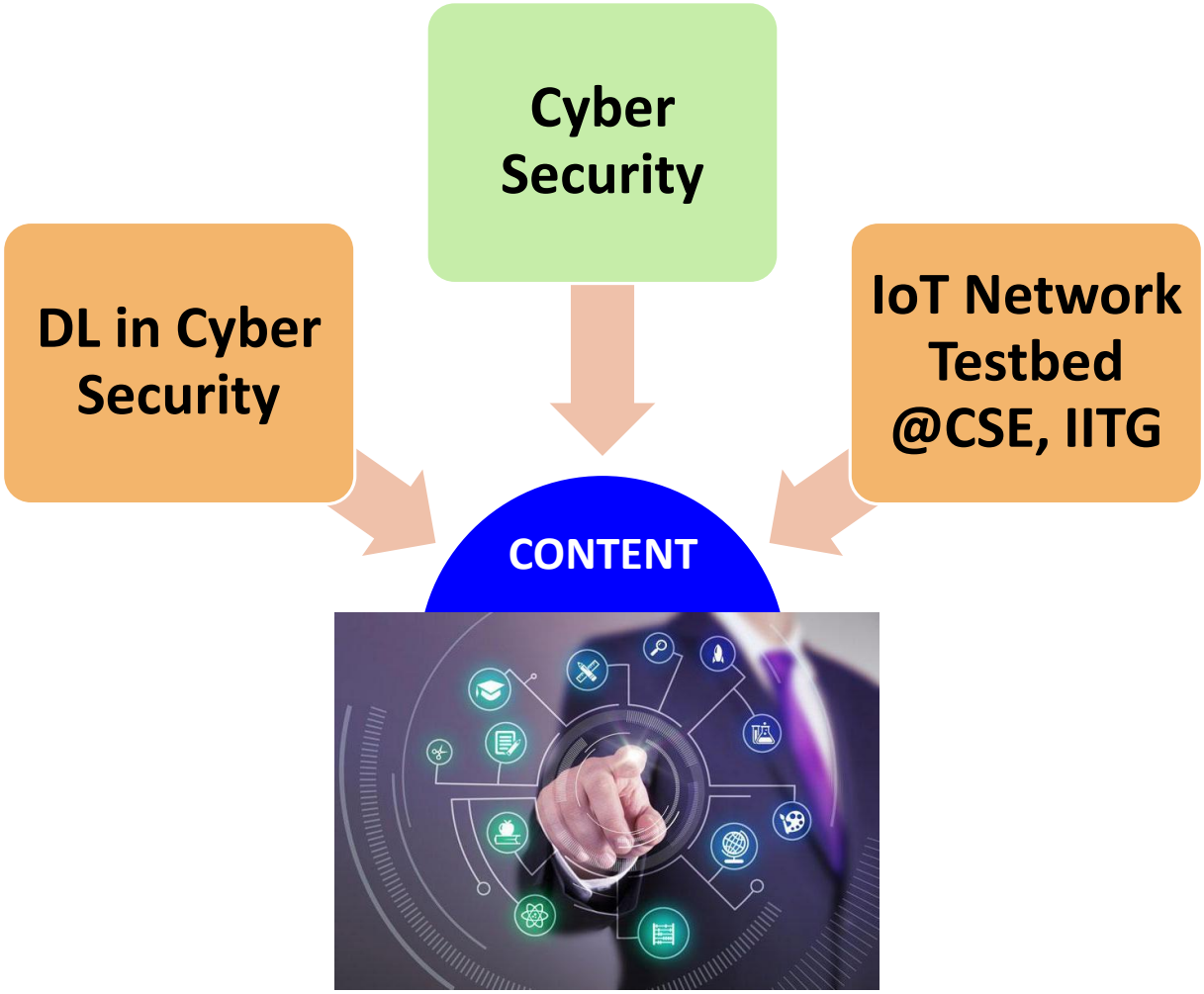
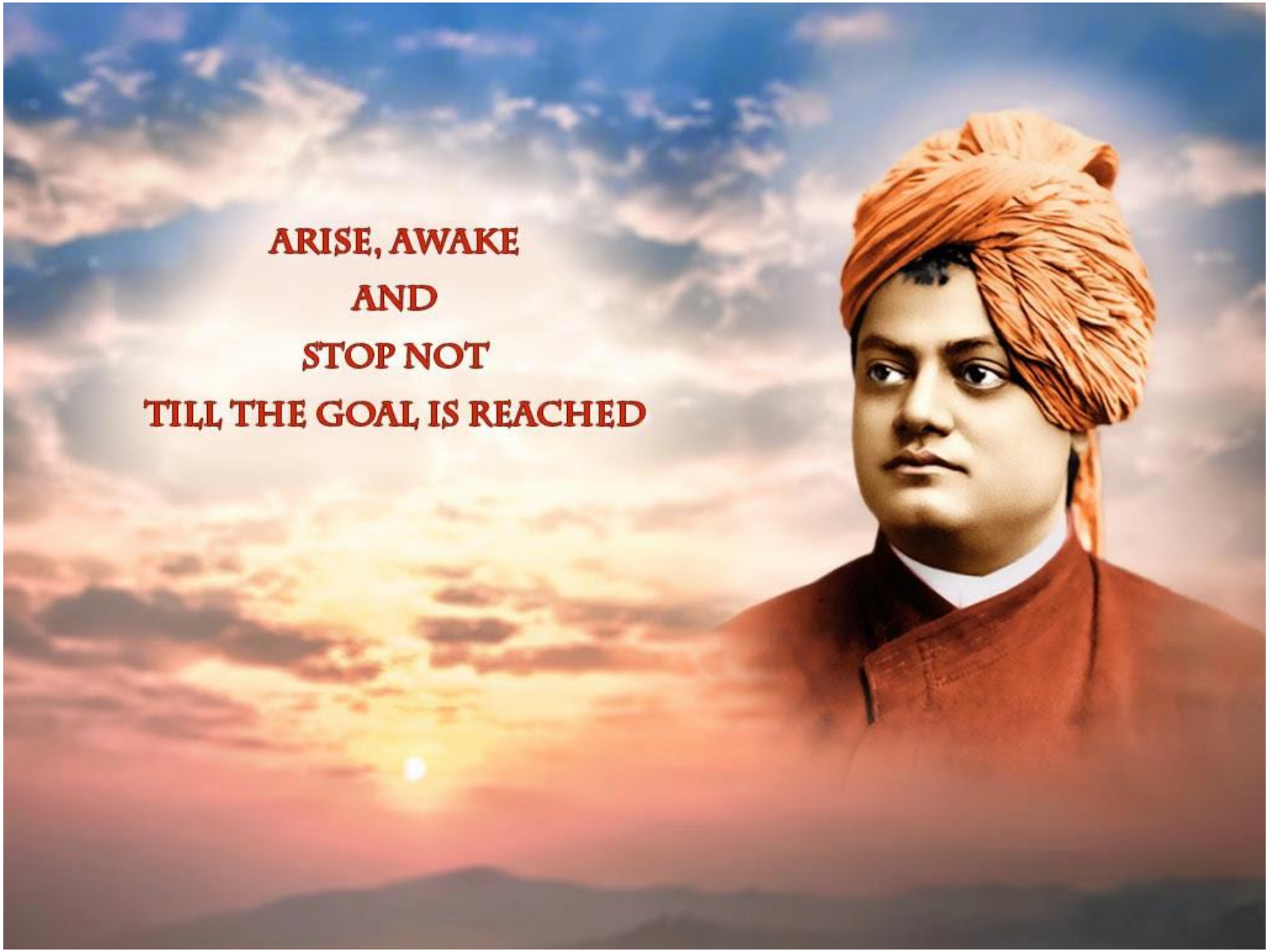


Image Source: https://www.sundayobserver.lk/2017/09/17/features/smart-revolution-role-education-creating-smart-people-smarter-world

National Youth Day



Learning in Life

श्रद्धावान् लभते ज्ञानं
(śraddhāvān labhate jñānam)

"To me the **very essence of education is concentration of mind**, not the collecting of facts."

"**Training of mind should be a student's highest priority**, and not simply the accumulation, the memorizing, and the repeating of facts".

"What a man 'learns' is really what he 'discovers' by **taking the cover off his own soul**, which is a mine of infinite knowledge".

What is Education? Is it book-learning? No. Is it diverse knowledge? Not even that. Training by which the **current and expression of will are brought under control** and become fruitful is called education.

Every boy **should be trained to practice absolute *brahmacharya,*** and then *sraddha*, faith, will come.

What is Cyber Security ?

Cyber Security: a definition

- Cyber security is how individuals and organisations reduce the risk of cyber attacks
- Cyber security's core function is to protect
  - the **devices** we all use (smartphones, laptops, tablets and computers), and
  - the **services** we access both online and offline from theft or damage.

- *National Cyber Security Research Council* (NCSRC), India, aimed to safeguard the Nation from the current threats in the Cyberspace towards National Security.
- "I dream of a digital India where cyber security becomes an integral part of our National Security" – Prime Minister of India

Image Source: https://www.google.com/

Types of Cyber Threats & Attacks

- **Malware** (malicious software) - all kinds of software created by bad actors to damage devices, systems, and networks.
- **Phishing -** act of sending infected emails or messages disguished as legitimate to trick victims into giving personal and valuable data or downloading malware.
- **DoS attack -** attackers use this technique to flood networks and servers with traffic, causing resource drain, and making them unavailable.
- **Data breach -** an unauthorized user gains access to valuable and confidential data such as user and credit card information.
- **Insider threats -** an attack caused by internal entity (person/device) employed by the company.
- **MITM** attack - when a perpetrator positions himself in a conversation between two legitimate parties — either to eavesdrop or to impersonate one of the parties, making it appear as if a normal exchange of information is underway.
- **Advanced Persistent Threats (APT) -** attacks capable of evading traditional defensive and perimeter security tools due to their stealthy nature.

National Critical Information Infrastructure

Banking, Financial Service, Insurance

Healthcare

Transport

Oil and Gas

Telecom

Power and Energy

Government Sector

Image source: https://nciipc.gov.in/

Few Recent Cyber Attacks

1

On 26 December 2024, Japan Airlines (JAL) fell victim to a massive Cyber Attack. Suspected DDoS Attack.

https://cybersecureindia.in/japan-airlines-hit-by-cyberattack-details-magnitude-and-impact/

2

Digital Arrest

'ডিজিটাল অ্যারেস্ট' হলেন বর্ধমানের 'সাধুবাবা'! কিছু মিলবে না বুঝে প্রণাম করে রণে ভঙ্গ দিল 'পুলিশ'

শেষ আপডেট: ০৮ জানুয়ারি ২০২৫ ১৪:০২

Fraudsters posing as law enforcement officers or cybercrime agents threatened victims with legal action unless they paid.

https://www.anandabazar.com/topic/digital-arrest

3

Home / Companies / News / BSNL data breach exposes 278 GB of sensitive telecom info, twice in 6 mts

BSNL data breach exposes 278 GB of sensitive telecom info, twice in 6 mts

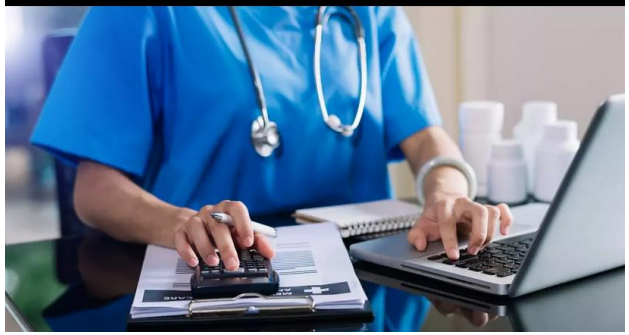
A threat actor has claimed to have obtained sensitive data, which includes international mobile subscriber identity numbers, SIM card specifics, home location register data, and security keys

BSNL suffered a security data breach on 26 June 2024

https://www.business-standard.com/companies/news/bsnl-data-breach-exposes-278-gb-of-sensitive-telecom-info-twice-in-6-mts-124062600314_1.html

4

Change Healthcare cyberattack causes dire billing crisis



UnitedHealth of USA, which owns Change Health, reportedly paid $22 million in ransom. Attack happened on 21 February 2024.

https://securityintelligence.com/news/change-healthcare-cyberattack-billing-crisis/

Few Recent Cyber Fraud

কিউআর কোডে ডিজিটাল ফাঁদ! সাদাকালো সঙ্কেত না বুঝে স্ক্যান করলেই বিপদ, কী ভাবে কী কী হতে পারে

কিউআর কোড। বর্গাকার সাদাকালো সাঙ্কেতিক একটি ছবি। অনলাইনে আর্থিক লেনদেনের জন্য যা অপরিহার্য। তারই ফাঁকফোকরে গজিয়ে উঠছে প্রতারণার ফাঁদ। সময় থাকতে যা চিনে নেওয়া জরুরি। আজ প্রথম কিস্তি।

5

https://www.anandabazar.com/west-bengal/how-people-are-misusing-qr-code-for-digital-crime-and-how-to-avoid-it-dgtls/cid/1573461

7

CrowdStrike IT outage affected 8.5 million Windows devices, Microsoft says

20 July 2024

Share Save

Joe Tidy
Cyber correspondent, BBC News

On 20 July 2024, global network outage linked to CrowdStrike was caused by a flawed update to a cloud-based security software of CrowdStrike.

https://www.bbc.com/news/articles/cpe3zgznwjno

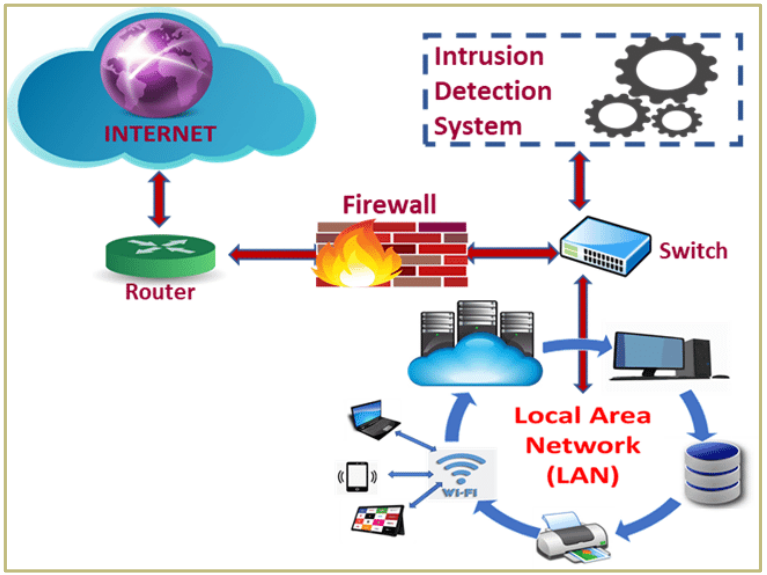
6

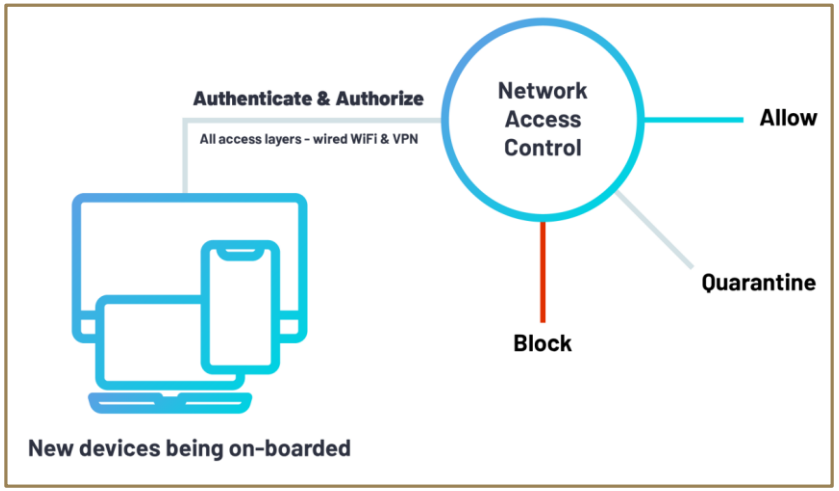
https://www.youtube.com/watch?v=aovyByF2lBw ; https://decentro.tech/blog/upi-frauds/

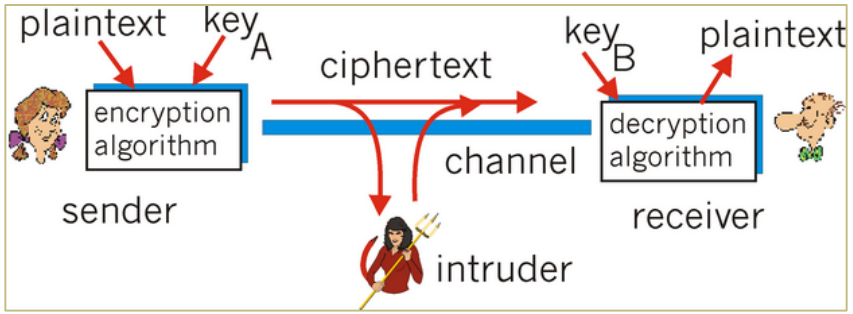
The " **Cyber Swachhta Kendra** " (Botnet Cleaning and Malware Analysis Centre) by MeitY, Government of India.

https://dot.gov.in/ ; https://www.csk.gov.in/

Common Defense Strategies







Cryptography

Image source: https://www.linkedin.com/pulse/what-nac-network-access-control-ronen-taieb
Image source: https://www.indiamart.com/proddetail/antivirus-software-11021229112.html
Image source: G. Karatas et. al., "Increasing the Performance of Machine Learning-Based IDSs on an Imbalanced and Up-to-Date Dataset," in *IEEE Access*, vol. 8, pp. 32150-32162, 2020.

#BreakTheChain

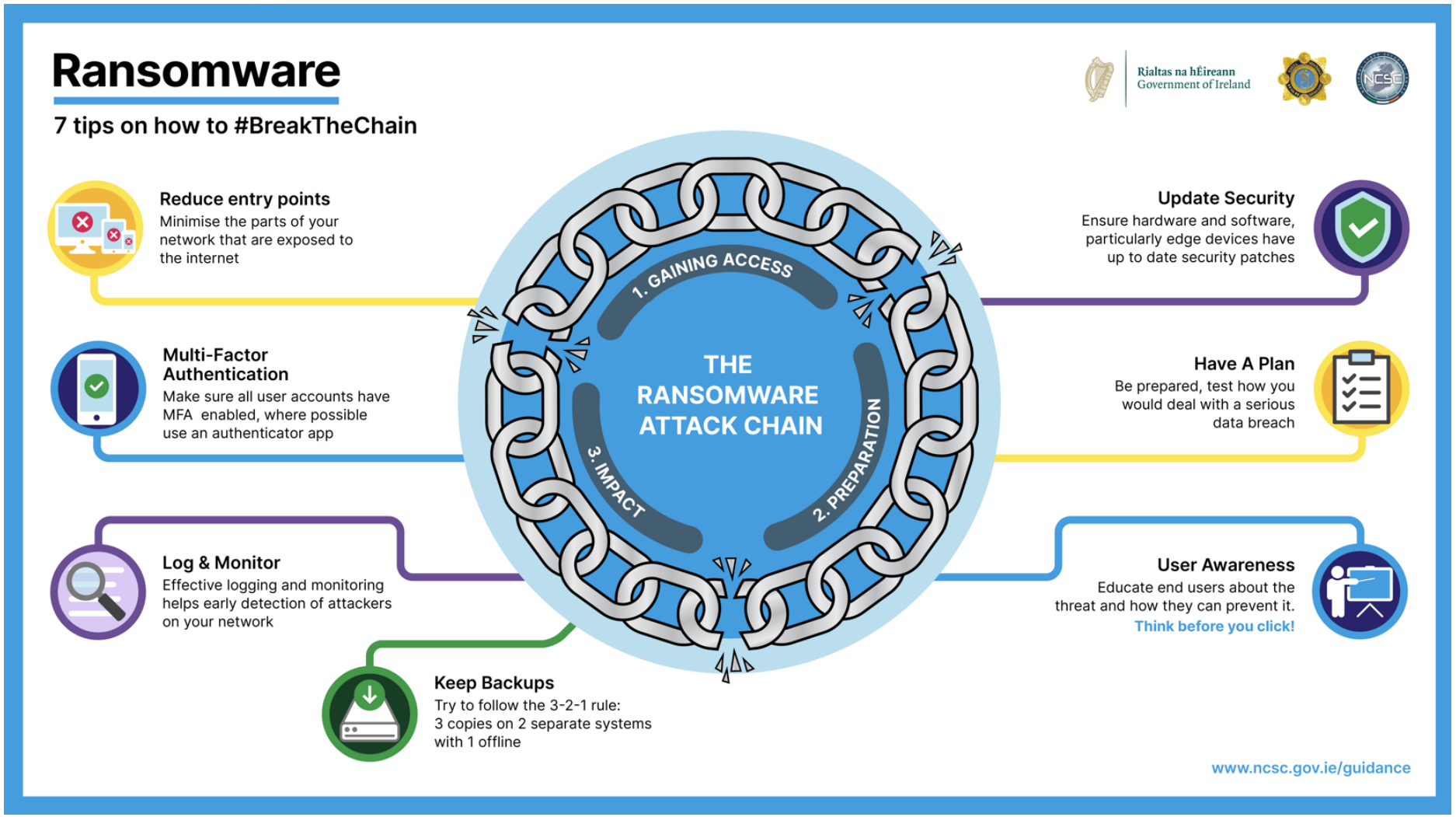


Image source: https://ncsc.gov.ie/pdfs/NCSC_Quick_Guide_Ransomware.pdf

The Ransomware Attack Chain

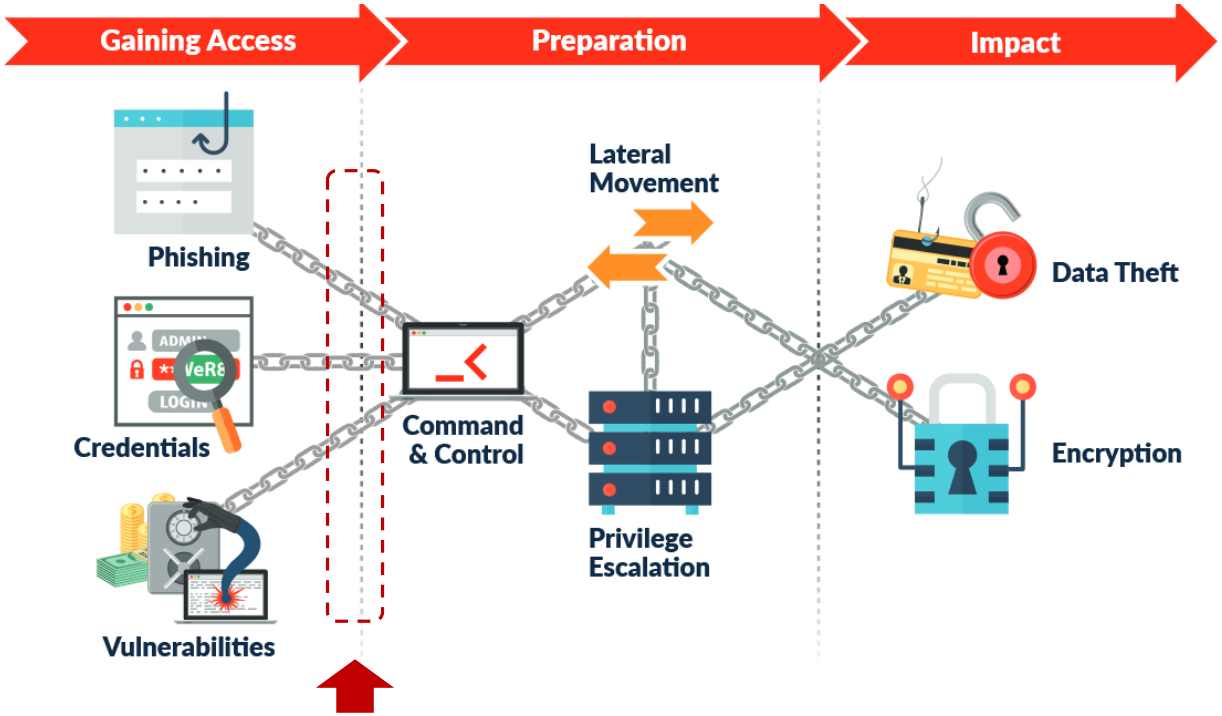


Image source: https://ncsc.gov.ie/pdfs/NCSC_Quick_Guide_Ransomware.pdf

Where DL can play the role?

- In order to carry out a ransomware attack, the attacker **needs to gain access** to your Network or System.
- *Example*: The attacker sends an **email** which tricks a user in to downloading a malicious file or clicking on a malicious link. This will result in some form of malware being downloaded, usually a backdoor, which will provide the attacker with initial access to the system.

Why DL in Cyber Security ?

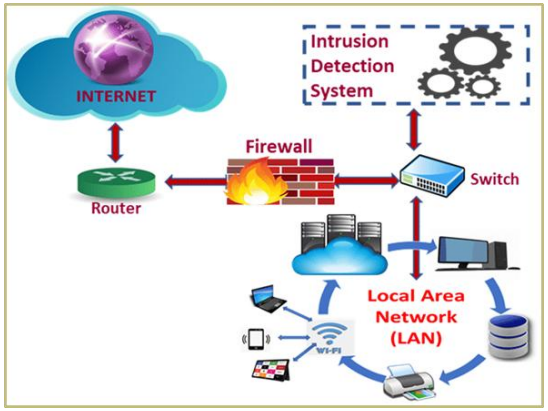
- Existing traditional cyber security solutions fail to address the growing dynamics of modern cyberattacks
  - ✓ detecting new threats
  - ✓ analysing complex vectors and events
  - ✓ ability to scale to the sheer volume of attacks

❖ Applying deep learning (DL) in cyber security can eliminate these problems

What makes the DL algorithm suitable in Cyber Security solutions ?

- DL models are able to process massive volumes of raw data that are used to automatically train the cyber security system.
- DL is able to accurately identify highly complex patterns from large data sets
- DL has its ability to proactively identify and stop attacks before they happen.
- DL algorithms can adjust themselves to data properties they are trained on
- DL continues to evolve and learn over time to pre-emptively recognize threats it has not seen before. So, DL is effective against unknown or zero-day threats.

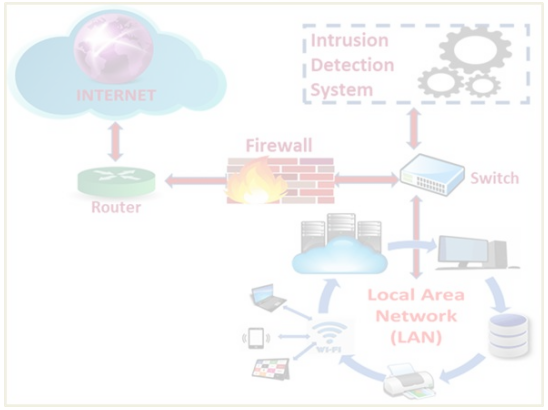
Intrusion Detection System (IDS)

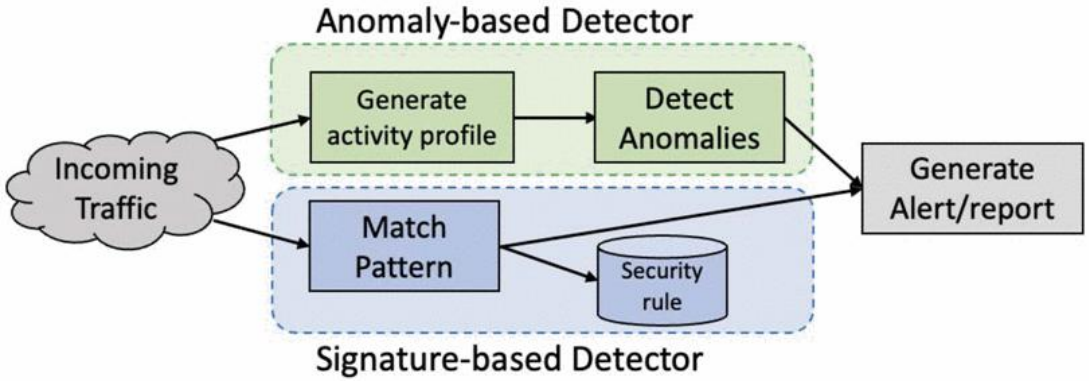


- Network-based IDS: NIDS act as network monitoring devices deployed at strategic points within a computer network.
  - ✓ capture and **analyse network traffic data**

- Host-based IDS: HIDS function as software agents deployed directly on the operating system of the host device itself.
  - ✓ monitor and **analyse activity** occurring on the host device.

Intrusion Detection System (IDS)





- Network-based IDS: NIDS act as network monitoring devices deployed at strategic points within a computer network.
  - ✓ capture and **analyse network traffic data**
- Host-based IDS: HIDS function as software agents deployed directly on the operating system of the host device itself.
  - ✓ monitor and **analyse activity** occurring on the host device.

- Signature-based IDS relies on a predefined database of attack signatures to identify malicious activity.
- Anomaly-based IDS identify deviations from normal behaviour in a network or system to detect computer and network intrusions.

DL Methods used in Cyber Security

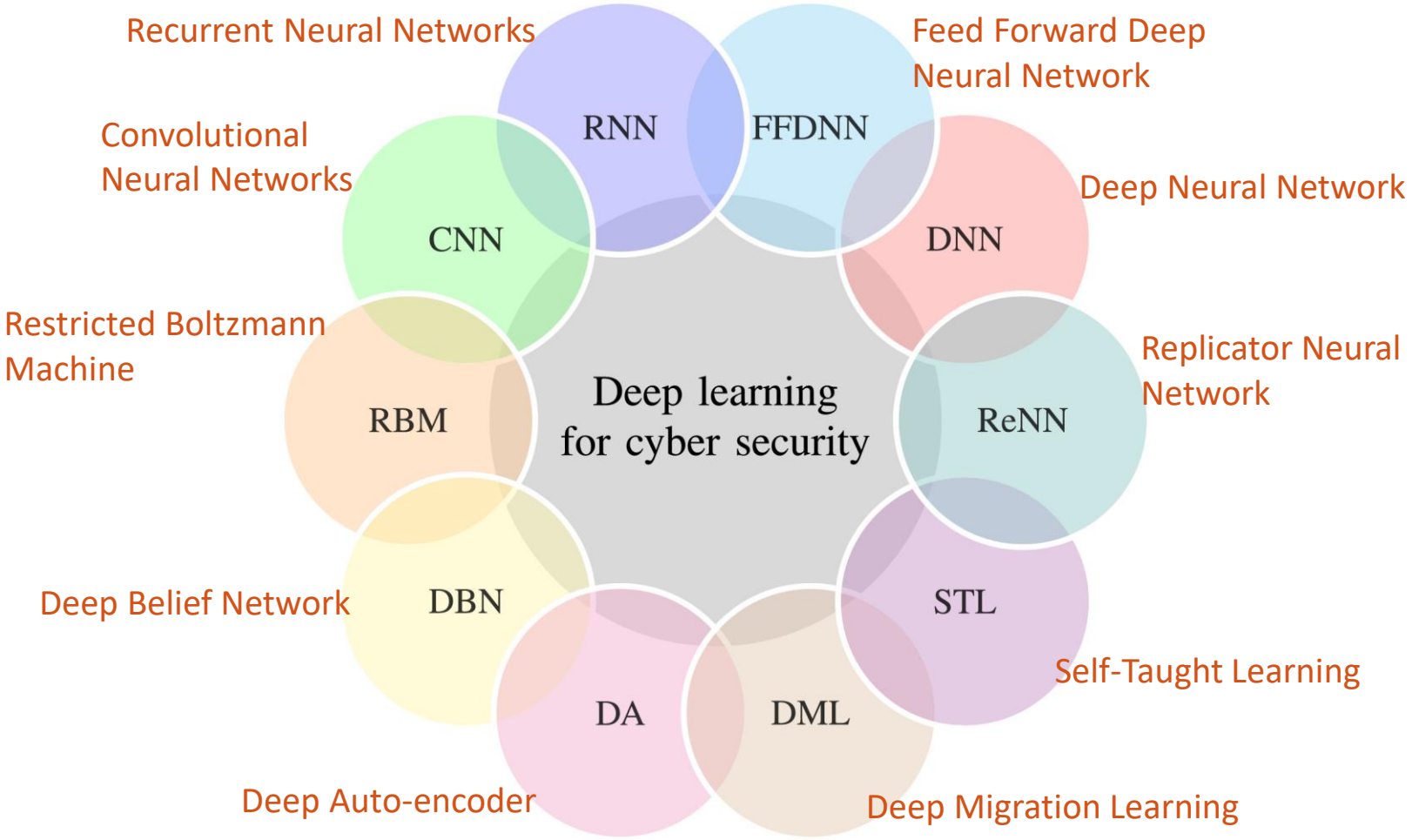
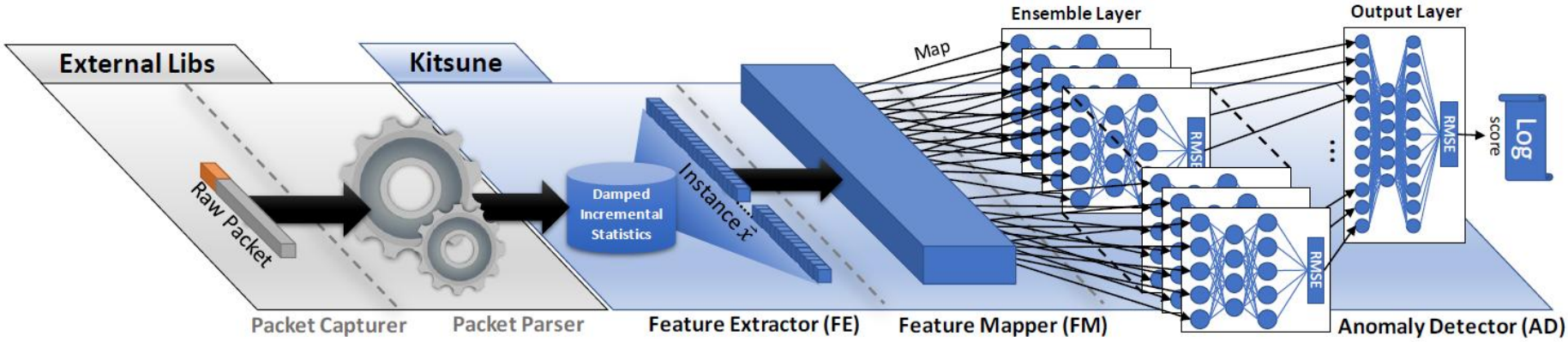


Image source: M. A. Ferrag *et. al.*, "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study", in *Journal of Information Security and Applications*, vol. 50, 2020.

Kitsune NIDS Architecture



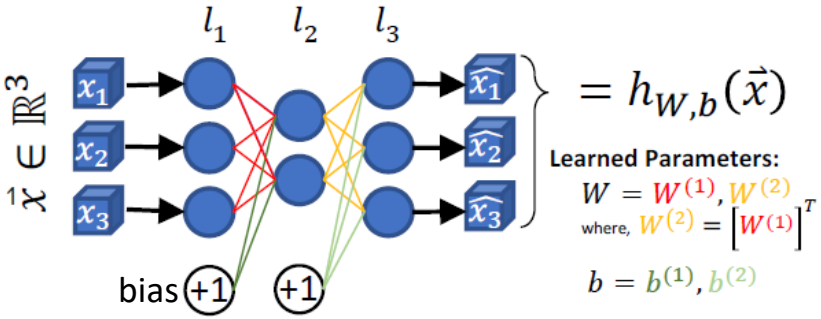


Fig: An example autoencoder with one compression layer, which reconstructs instances with three features.

Let ϕ be the anomaly threshold, with an initial value of -1, and let $\beta \in [1, \infty)$ be some given sensitivity parameter. One can apply an autoencoder to the task of anomaly detection by performing the following steps:

1) **Training Phase:** Train an autoencoder on clean (normal) data. For each instance x_i in the training set X:
   a) Execute: $s = \text{RMSE}(\vec{x}, h_\theta(\vec{x}))$
   b) Update: if$(s \geq \phi)$ then $\phi \leftarrow s$
   c) Train: Update θ by learning from x_i
2) **Execution Phase:**
   When an unseen instance $\vec{x}$ arrives:
   a) Execute: $s = \text{RMSE}(\vec{x}, h_\theta(\vec{x}))$
   b) Verdict: if$(s \geq \phi\beta)$ then *Alert*

Image source: Y. Mirsky *et. al.*, "Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection", in Proc. of 25th *NDSS*, 18-21 February 2018, San Diego, CA, USA.

Kitsune NIDS Performance

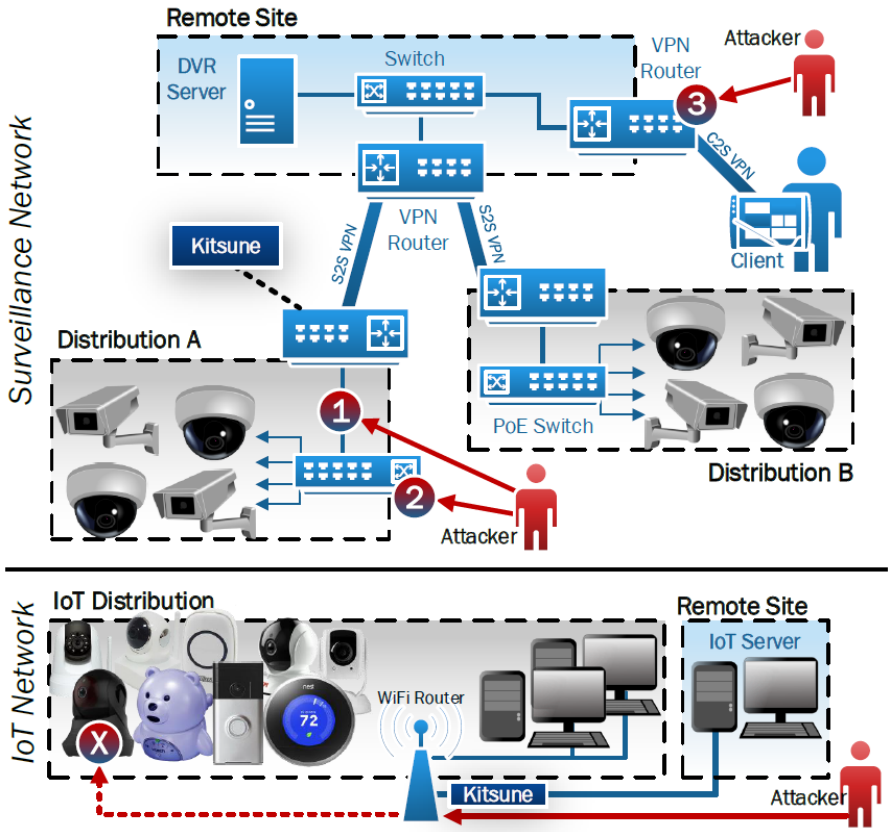


Fig: The network topologies used in the experiments

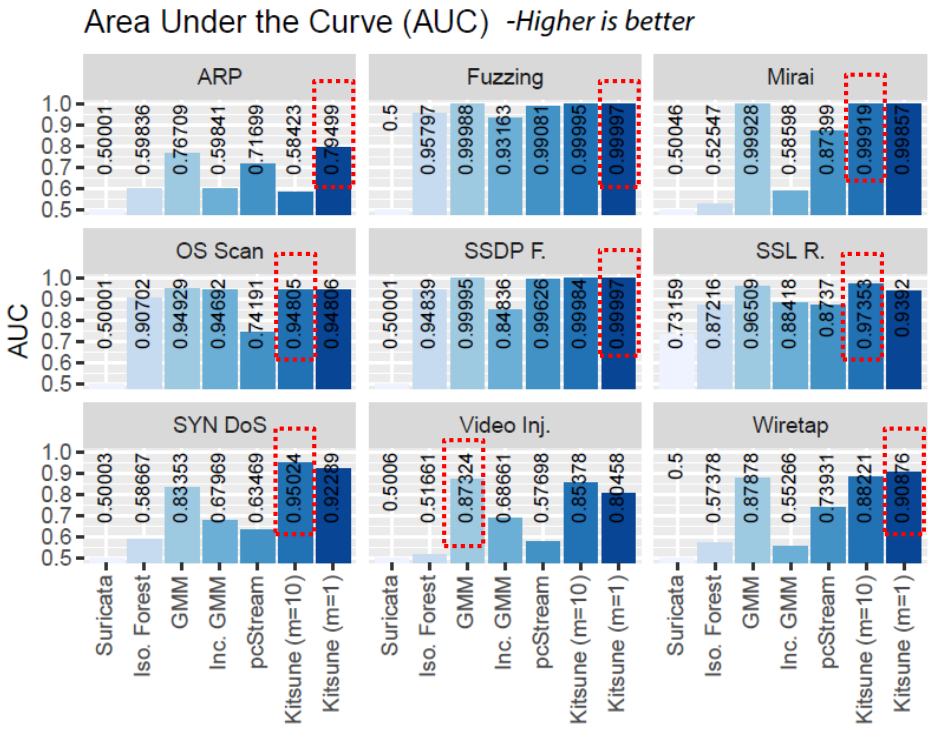
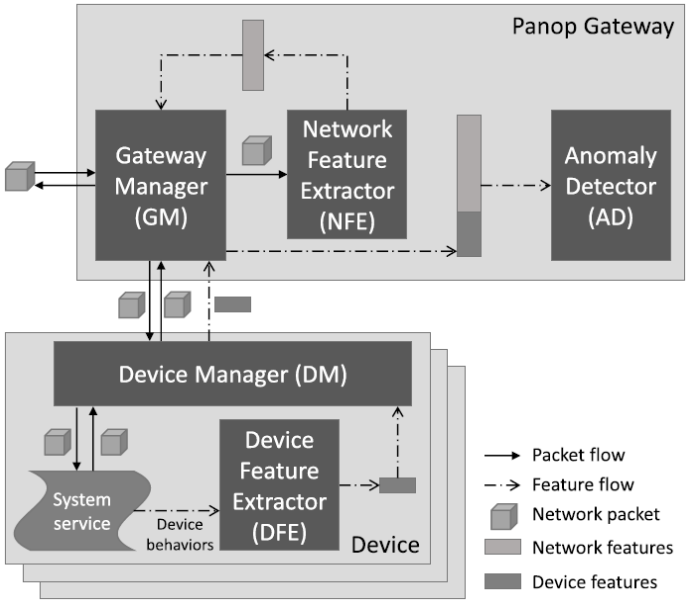


Fig: The experimental results for all algorithms on each of the datasets

Image source: Y. Mirsky *et. al.*, "Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection", in Proc. of 25th *NDSS*, 18-21 February 2018, San Diego, CA, USA.

Panop: ANN-based Distributed NIDS



Feature Types: System call sequence, program branch execution & network traffic.

Uses ensemble of shallow AEs and one LSTM

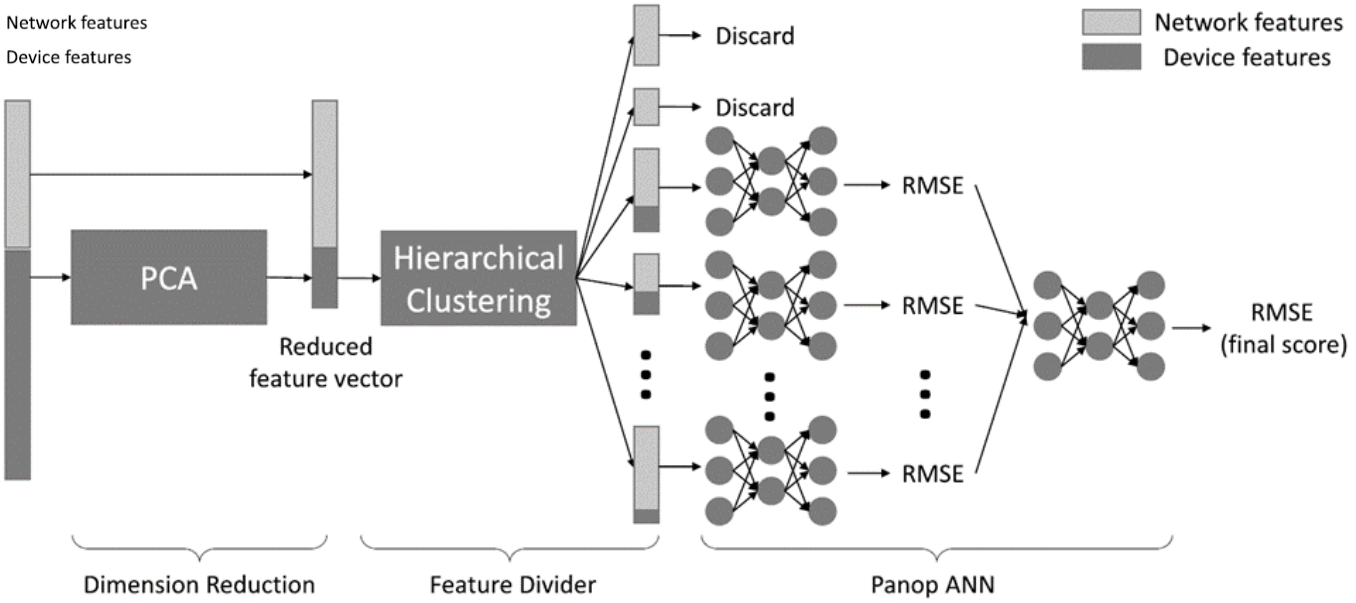


Image source: H. Kim et al., "Panop: Mimicry-Resistant ANN-Based Distributed NIDS for IoT Networks," in *IEEE Access*, vol. 9, pp. 111853-111864, 2021.

Detection Accuracy in Panop

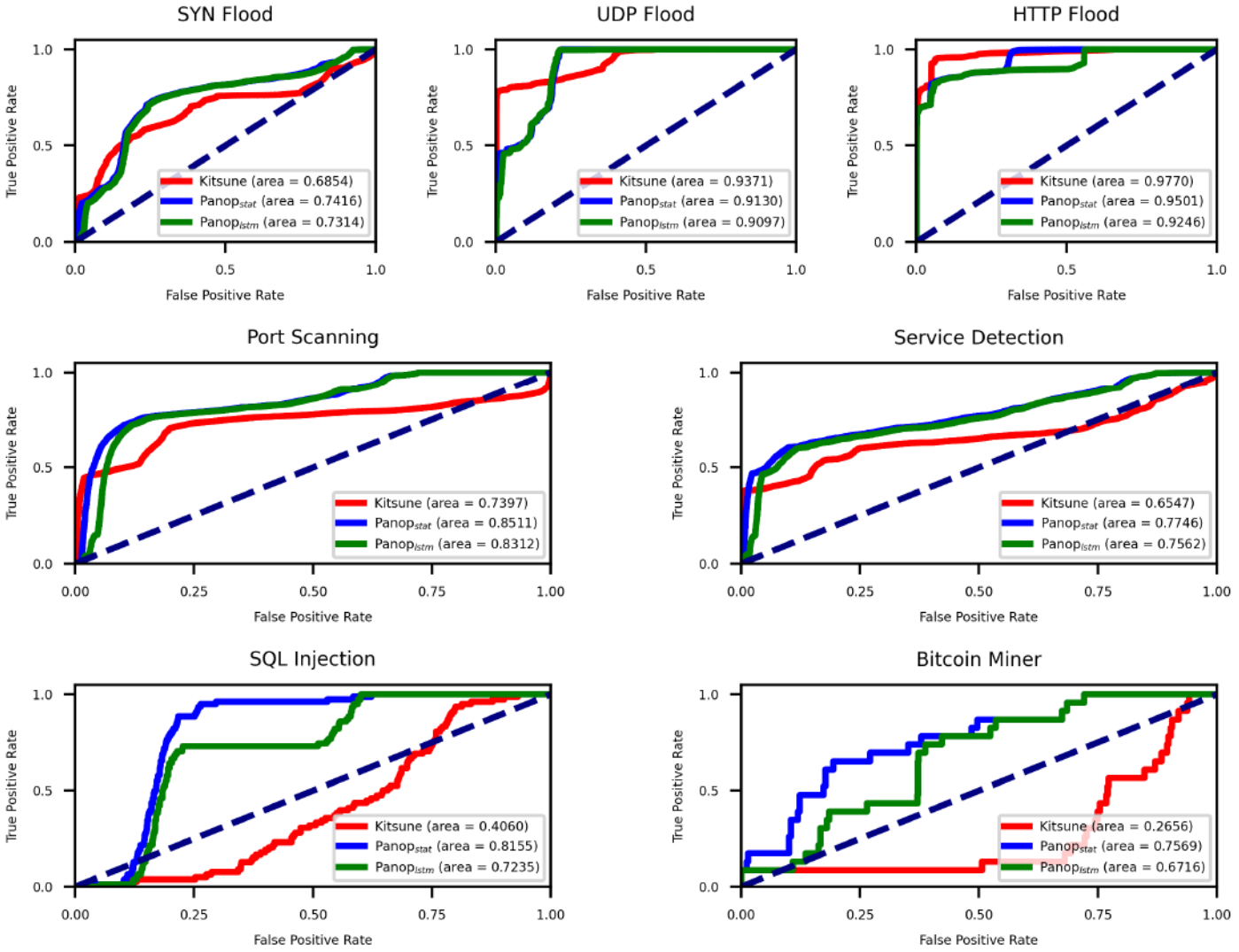
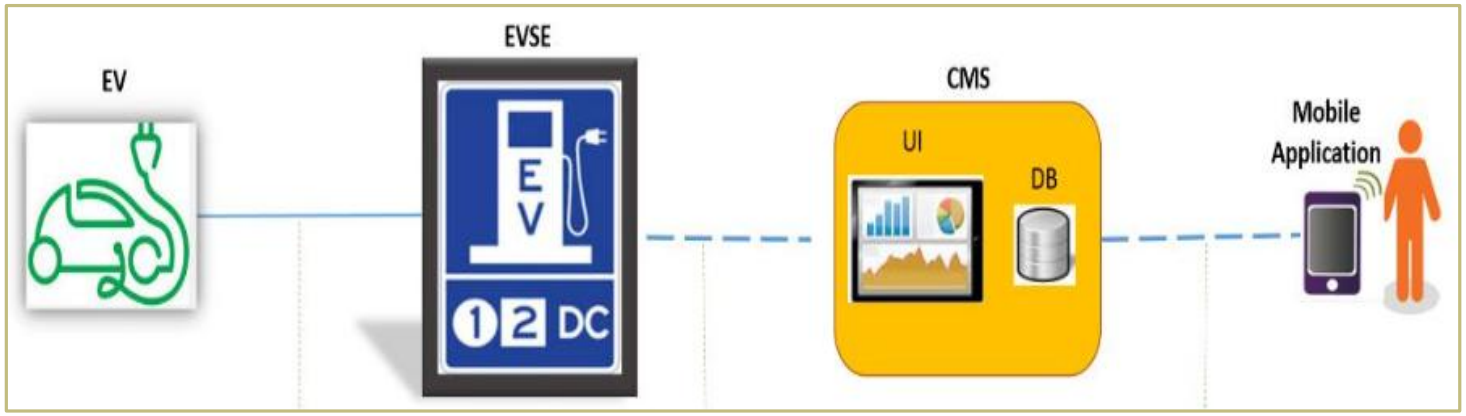


Image source: H. Kim et al., "Panop: Mimicry-Resistant ANN-Based Distributed NIDS for IoT Networks," in *IEEE Access*, vol. 9, pp. 111853-111864, 2021.

FDIA Detection in EV Charging

Fig: Block diagram of EV Charging Infrastructure



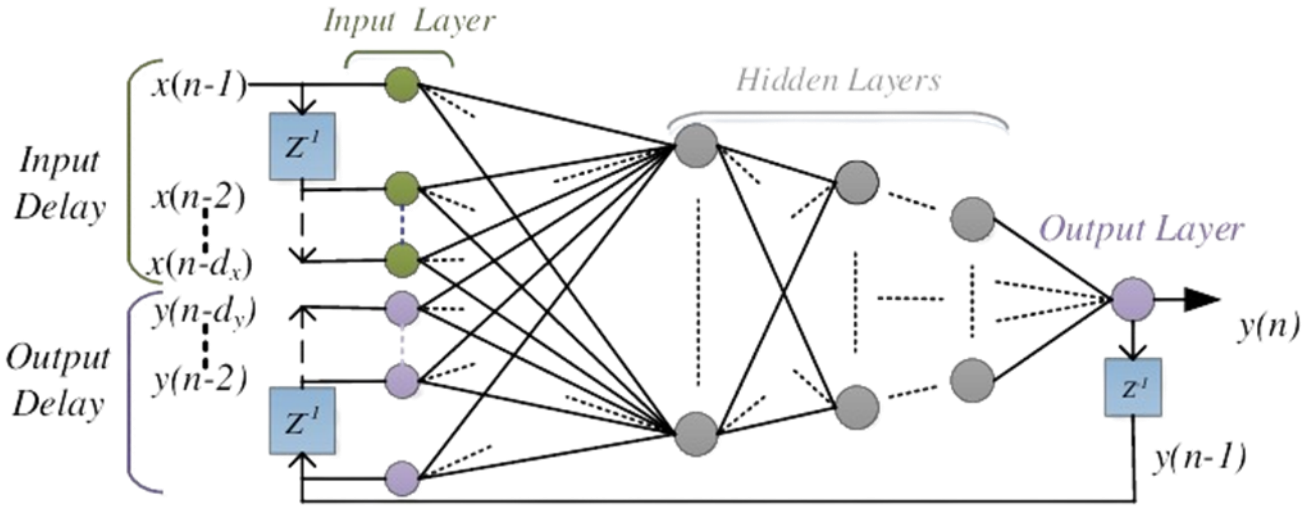
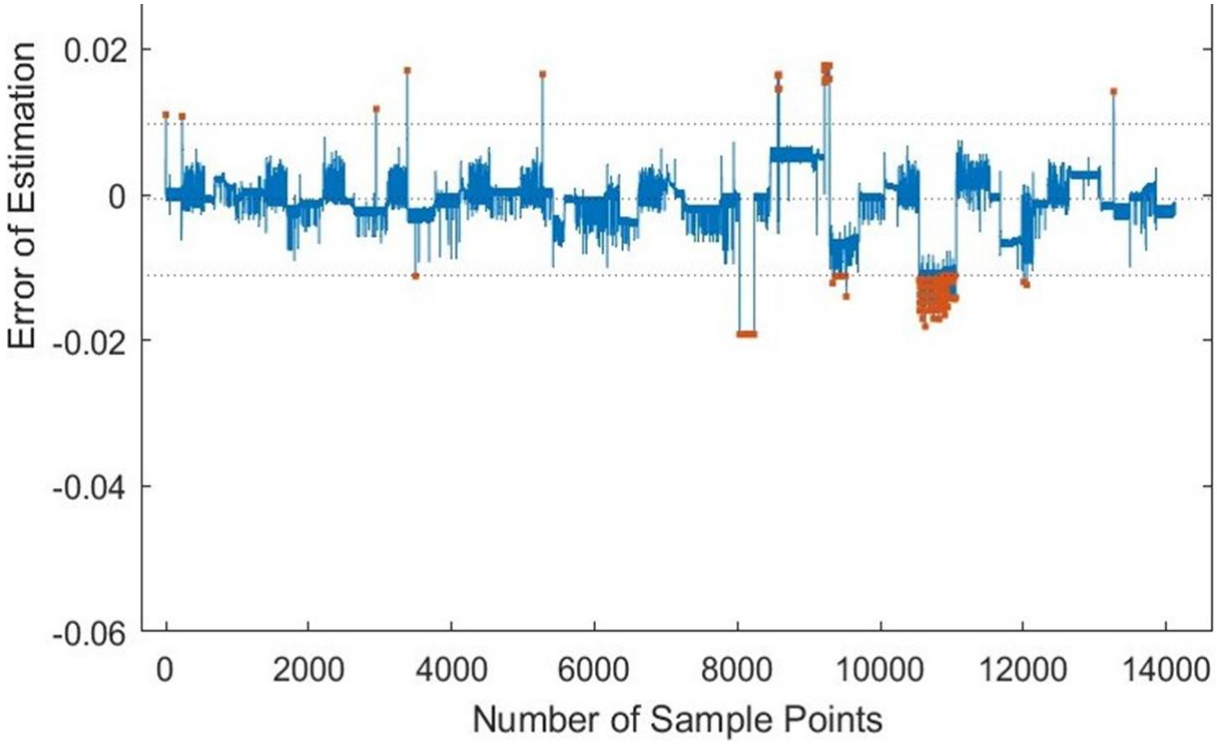


Fig: General Structure of NARX NN Architecture

Image source: Research Work Going On

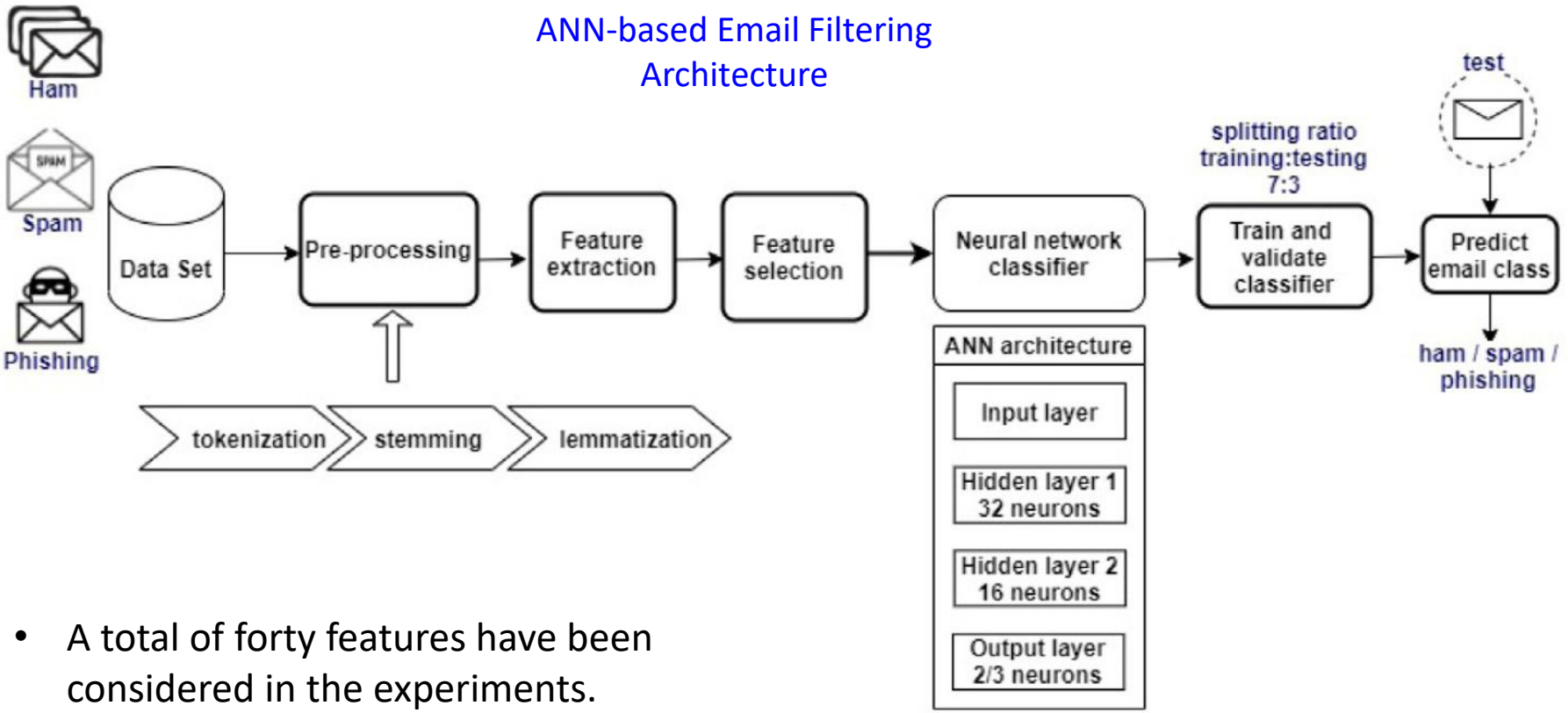
FDIA Detection in EV Charging



Parameter	Value
Accuracy	99.4 %
Recall	97.5 %
Precision	74.5 %
F1-Score	84.4 %

Image source: Research Going On

Email Filtering – Ham/ Spam/ Phishing



- A total of forty features have been considered in the experiments.
- Features are of five categories: body-based, subject line-based, sender address-based, URL-based, and script-based.

Image source: S. Magdy et. al., "Efficient spam and phishing emails filtering based on deep learning", in *Computer Networks,* vol. 206, 2022.

Email Filtering – Ham/ Spam/ Phishing

Datasets Studied

Distribution of different classes in studied datasets.

Dataset	Ham	Spam	Phishing	Number of classes
SpamBase	2788	1813	–	2
CSDMC2010	2949	1378	–	2
Phishing_corpus DS1	2758	660	–	2
Phishing_corpus DS2	2758	–	2432	2
Phishing_corpus DS3	2758	660	2432	3

Average performance metrics of applying our ANN classifier to the SpamBase dataset.

Epochs	SpamBase dataset (all features)					
	Accuracy	Precision	Recall	F1-score	MCC	Loss
100	0.9905	0.9905	0.9905	0.9905	0.8537	0.0238
200	0.9957	0.9965	0.9965	0.9965	0.8427	0.0153
300	0.9950	0.9968	0.9968	0.9968	0.8519	0.0183

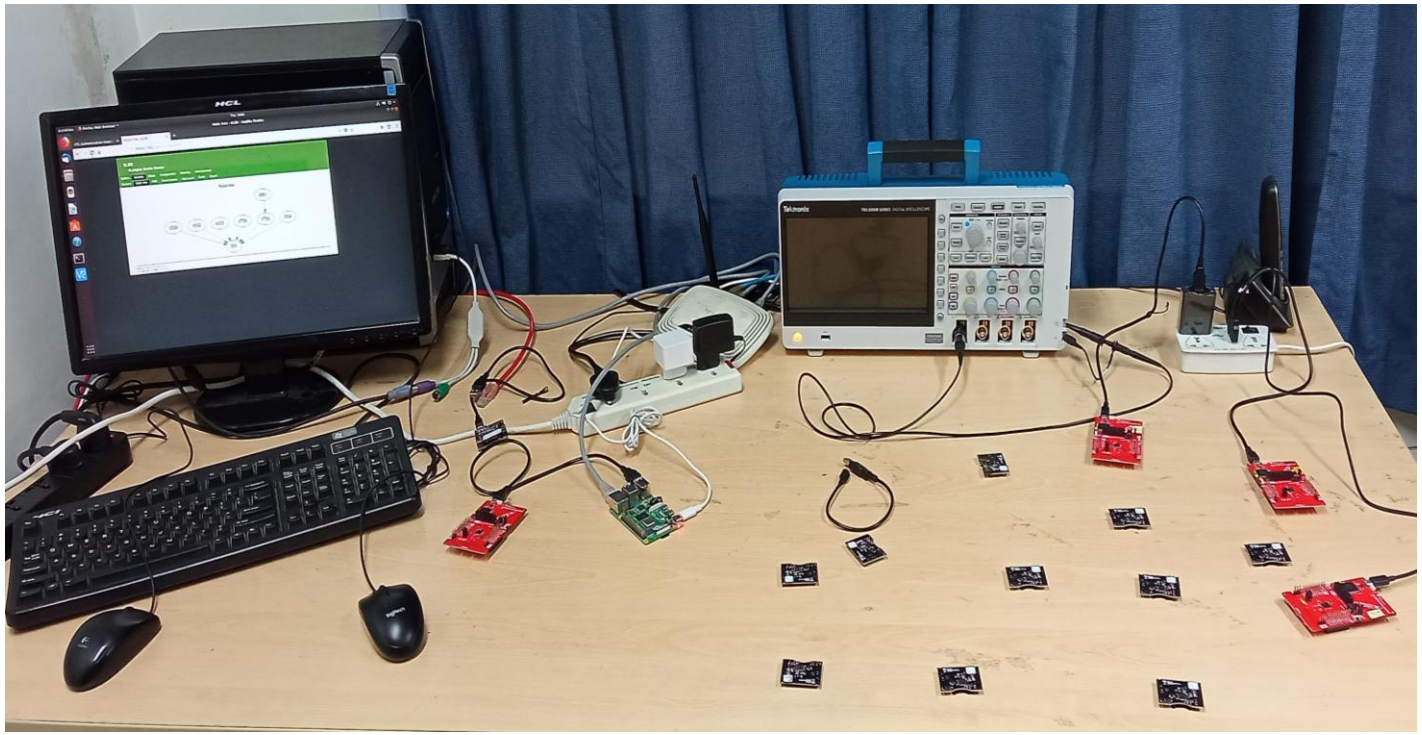
SpamBase Dataset

Phishing Dataset

DS	All features — 200 epochs						
	Acc.	Loss	Val. acc.	Val. loss	Prec.	Recall	F1-score
DS1	0.9963	0.0065	0.960	0.3895	0.9965	0.9965	0.9965
DS2	0.9982	0.0059	0.9859	0.2883	0.9965	0.9965	0.9965
DS3	0.9891	0.0377	0.9561	0.6283	0.9860	0.9831	0.9844

Image source: S. Magdy et. al., "Efficient spam and phishing emails filtering based on deep learning", in *Computer Networks,* vol. 206, 2022.

802.15.4 IoT Testbed @CSE, IITG



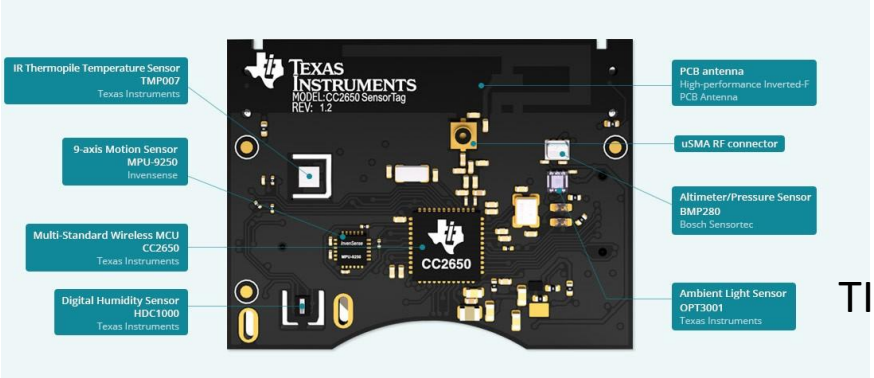
IoT Testbed

Raspberry Pi

TI CC2650 Launchpad

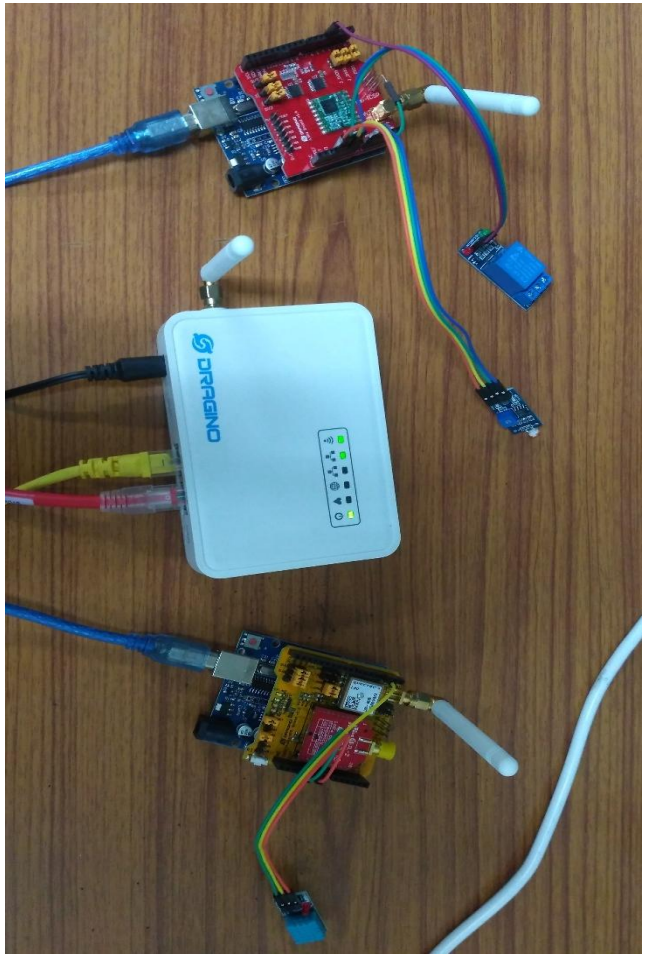
IEEE 802.15.4 TSCH Nodes

- 19 Launchpad
- 13 SensorTag



TI CC2650 SensorTag

LoRa IoT Testbed @CSE, IITG



LoRa IoT Testbed

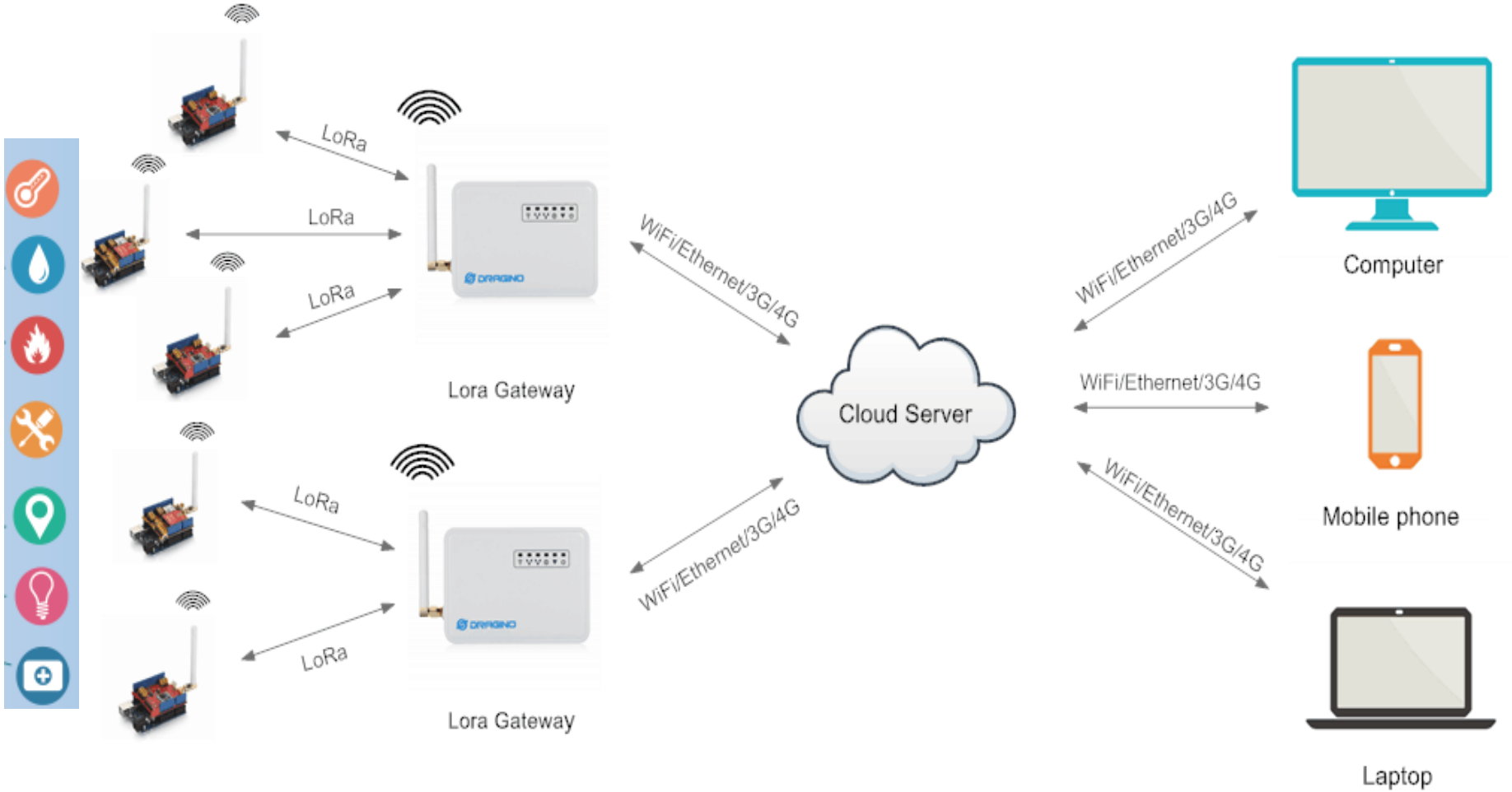
Single Channel
LoRa IoT Gateway

LoRa shield
upon R. Pi

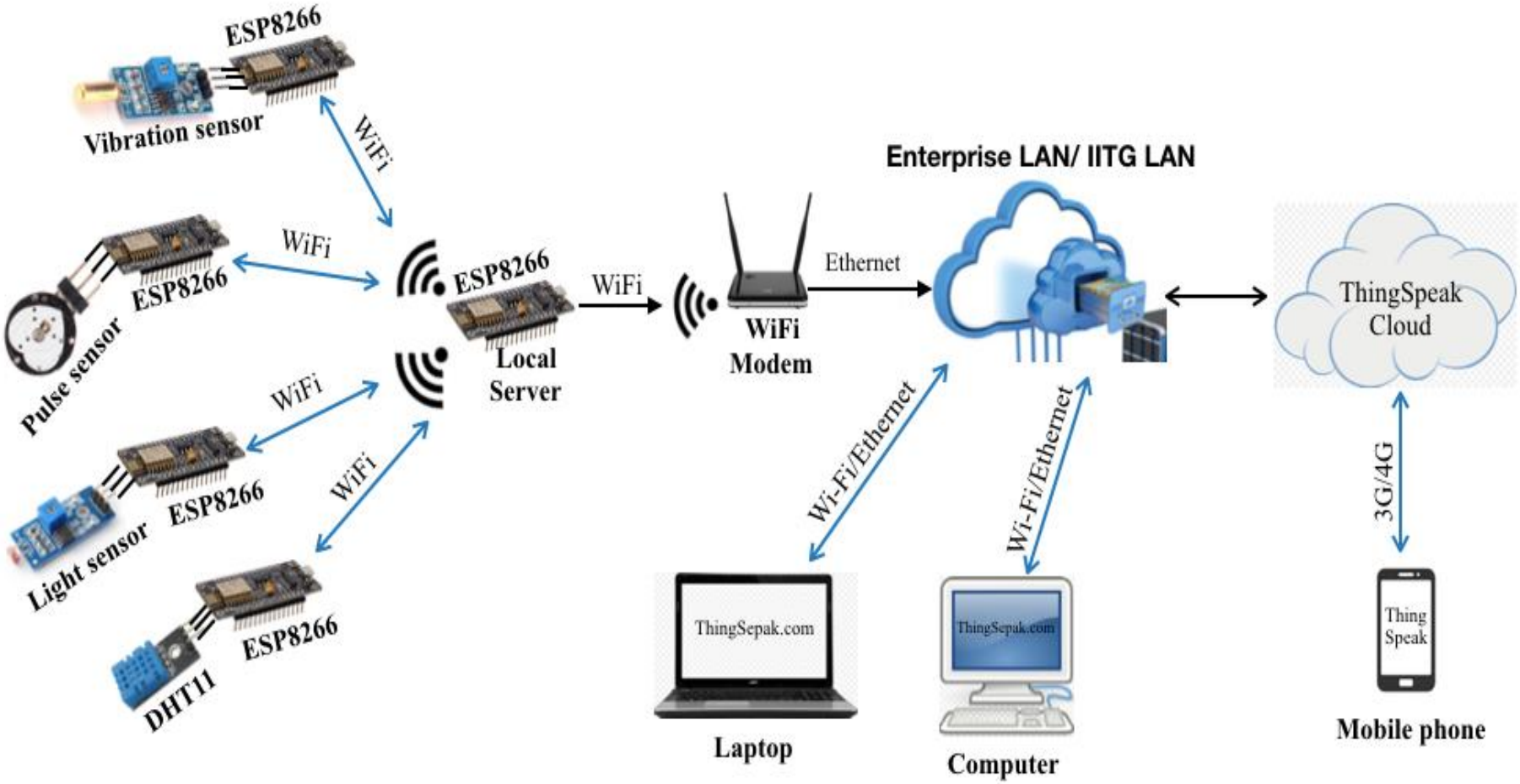
LoRa Shield
upon Arduino UNO

LoRa & GPS Shield
upon Arduino UNO

Long-range Monitoring & Control

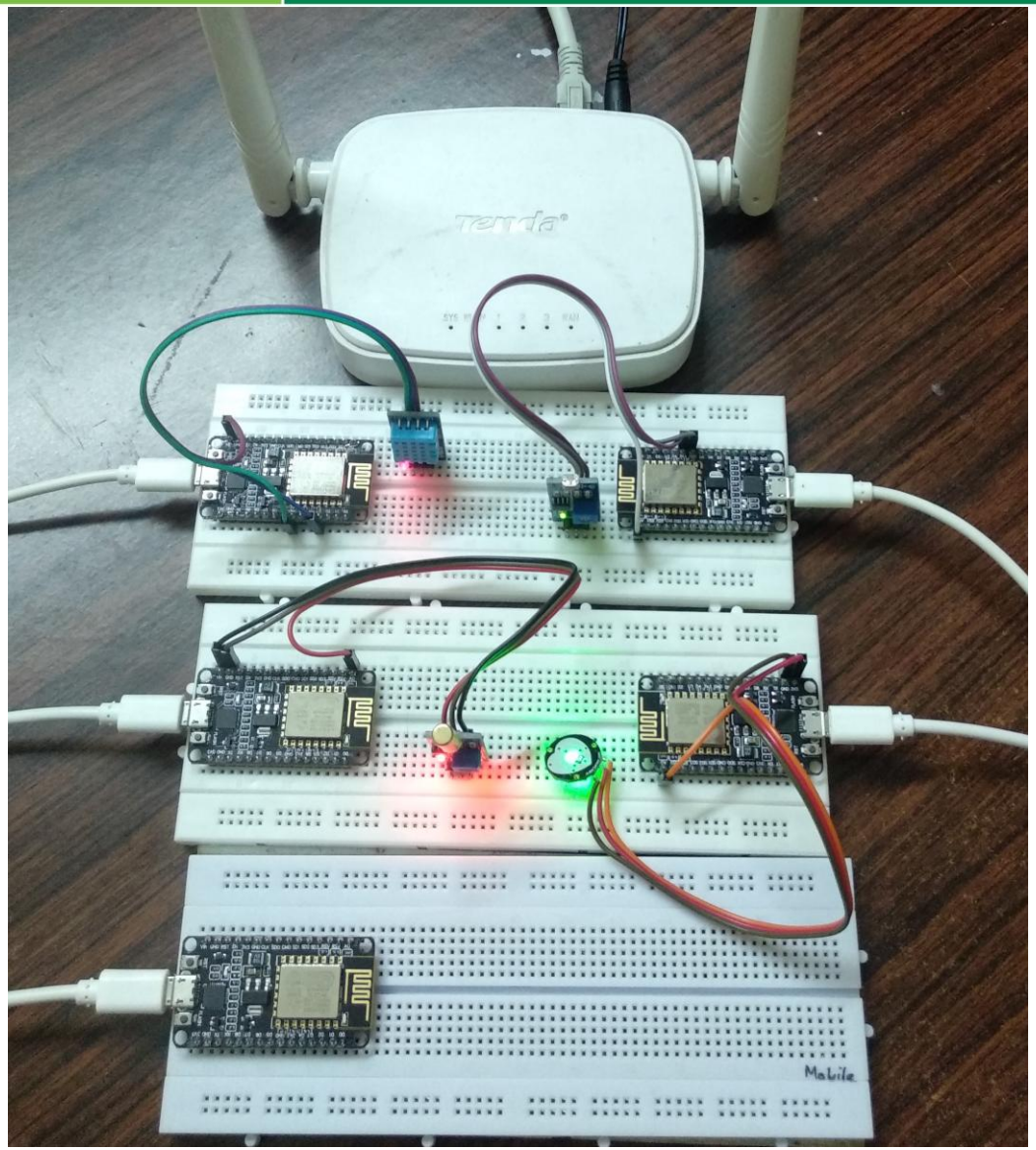


Basic IoT Setup @CSE, IITG



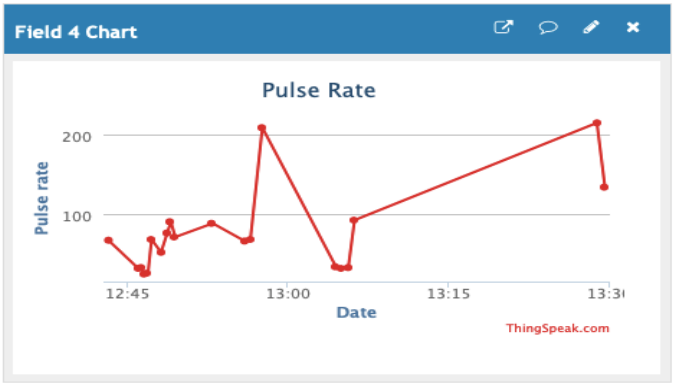
Smart Home Monitoring

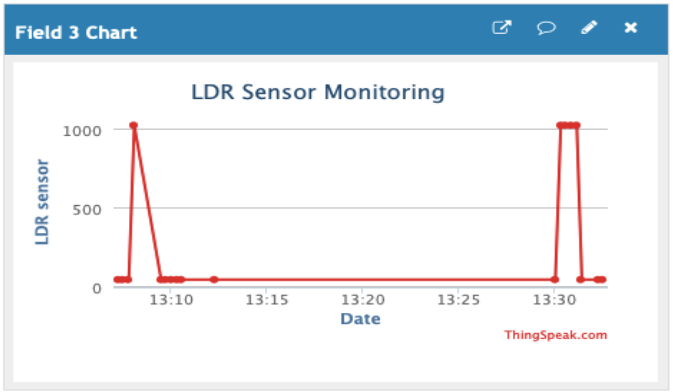
Basic IoT Setup @CSE, IITG



ThingSpeak Cloud Server

accessing from a Laptop/PC/Smartphone





Thank you

Questions and Discussion