BITS PILANI, HYDERABAD CAMPUS

BITS G450: Research Practice

M.E. CS-II Sem 2021-22

Submitted To:

Prof. Chittaranjan Hota

Submitted By:

Jainil Panchal 2021H1030071H

CERTIFICATE

This is to certify that the report entitled, "***Learning a fake node's (Sybil node's) behavior in an IoT-CPS system that could cause a Denial of Service (DoS) or could polarize a consensus using Deep learning techniques.***"

Submitted By:

Name	**Jainil Panchal**
BITS ID	**2021H1030071H**

In partial fulfillment of the requirement of BITS, G540 Research Practice embodies his work under my supervision.

Date: 04 / 05 / 2022

Submitted To:

Prof. Chittaranjan Hota
Dept. of Computer Science and Information Systems
BITS Pilani, Hyderabad Campus

INDEX

Abstract

A cyber-physical system (CPS) is a system that combines a physical system with the real world and controls applications in a computational system that interact over a network. Internet of Things (IoT) devices are getting more and more users daily with the developments in wireless sensor networks (WSN). However, increasing the number of connections improves the probability of a cyberattack in which attackers invade the network and perform cyber-physical attacks, remotely interrupting the CPS. These wireless sensor networks are susceptible to a wide range of malicious attacks. These attacks are intended to generate erroneous data or modify actual data transmitted across the network. One of these is a Sybil attack, in which a malicious node fabricates or obtains the identities of legitimate nodes. It can polarize the network. A system administrator must tackle these attacks to ensure seamless connectivity and coordination between IoT devices.

Introduction

Internet of Things (IoT)

The Internet of Things is a paradigm in which billions of everyday physical things in the actual world are integrated with sensing, actuating, networking, and computer processing capabilities. These objects - ”things” - are interconnected via a network and exchange data to achieve a specific goal. The term ”Internet of Things” is considered a misnomer since IoT devices connect to a network in which they are uniquely recognizable rather than the public internet. Connecting these devices and attaching sensors adds a degree of digital intelligence to otherwise dumb equipment, allowing them to relay real-time data without engaging a human.

We can classify these IoT devices/objects - as “things.”

- Trackable Object: Identifiable and aware of its whereabouts.
- Data Object: objects that generate data, such as sensors
- Interactive Object: An interactive item can communicate with its surroundings via monitoring environmental variables, modifying the environment, or both.
- Smart Object: an object having some processing capabilities to act on data received.

Cyber Physical System (CPS)

Cyber Physical Systems are sensor-based communication-enabled heterogeneous systems that combine control, communication, and interaction with the physical world via computing devices, sensors, and actuators. These devices are linked together on a large scale to automate processes due to event detection and specific decision procedures and connect the cyber and physical worlds. These are physical systems with engineering capabilities whose operations are controlled and monitored by a computing core. Examples of CPS Include Medical devices and systems, vehicles, Defense Systems, Robotic

Systems, and Factory Automation. As the application increases, new issues like digital signal processing, control system management, and communication & security. The most crucial of them will be security.

Relationships between IoT and CPS

Both IoT and CPS use sensor-based interactive technologies to increase the connections between cyberspace and the actual world. They did, however, come from separate backgrounds.

From an engineering and control standpoint, the CPS idea was born. A cyber-physical system is a collection of computer elements that work together to control physical entities. It occurs when software components are used to connect mechanical and electrical systems. They manage logistics and manufacturing systems autonomously using shared knowledge and information from processes.

The Internet of Things concept arose from networking and information technologies. The phrase ”Internet-of-Things” is used as a generic term to describe many features of the Internet and the Web’s expansion into the physical environment by deploying spatially distributed devices with embedded identification, sensing, and actuation capabilities.

IoT systems can still be networked together to manage a specific situation in a coordinated manner, at which point they are deemed to have grown to the level of a CPS. Because IoT involves monitoring objects in the real world, utilizing communication capabilities, and obtaining data needed to manage things that aren’t efficiently controlled now, it overlaps with CPS. Even though the Internet of Things was initially intended to focus on identification and monitoring technologies, it now encompasses a wide range of applications.

IoT-CPS

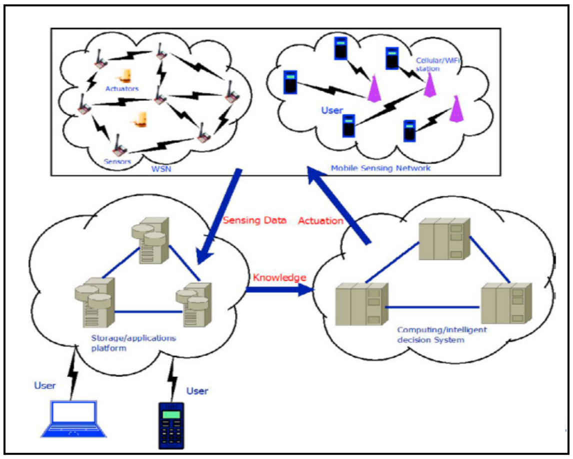


Fig.1 IoT-CPS Module

Fig.1 represents the actual flow of any IoT-CPS system, which includes sensing the data, gaining knowledge, and acting on the knowledge, i.e., actuation.

CPS/IoT Criteria

Any IoT-CPS system follows the following criteria.

1. The System has one or more components from the Logical, Physical, Transducing, and Human component categories.
2. Transmission, transformation, and storage are provided by integrating the physical elements. The information is contained in the logical components (data). Input, output, and processing functions are provided by the transducing components.
3. This function is responsible for connecting the system's logical and physical states.

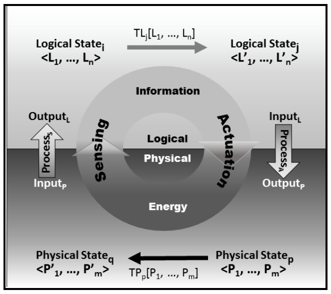

Logical State$_i$ $<L_1, \ldots, L_n>$
$TL_j[L_1, \ldots, L_n]$
Logical State$_j$ $<L'_1, \ldots, L'_n>$
Information
Output$_L$
Process$_S$
Input$_P$
Sensing
Logical
Physical
Actuation
Input$_L$
Process$_A$
Output$_P$
Energy
Physical State$_q$ $<P'_1, \ldots, P'_m>$
$TP_p[P_1, \ldots, P_m]$
Physical State$_p$ $<P_1, \ldots, P_m>$

Fig.2 Interactions model

The linked state transition diagram and connection are depicted in Fig2. After exchanging sensor data, the initial logical state transits to other logical states, resulting in actuator activation or transition from one physical state to another.

5C - Architecture

1. **Connection:** Data acquisition stage from devices or manufacturing systems such as ERP, manufacturing execution system
2. **Conversion:** the process of extracting valuable information from collected data.
3. **Cyber:** the stage at which information is sent to linked devices over a network.
4. **Cognition:** the step of optimizing division using learned information using machine learning methods.
5. **Configuration:** the process of transferring a decision from cyberspace to physical space.

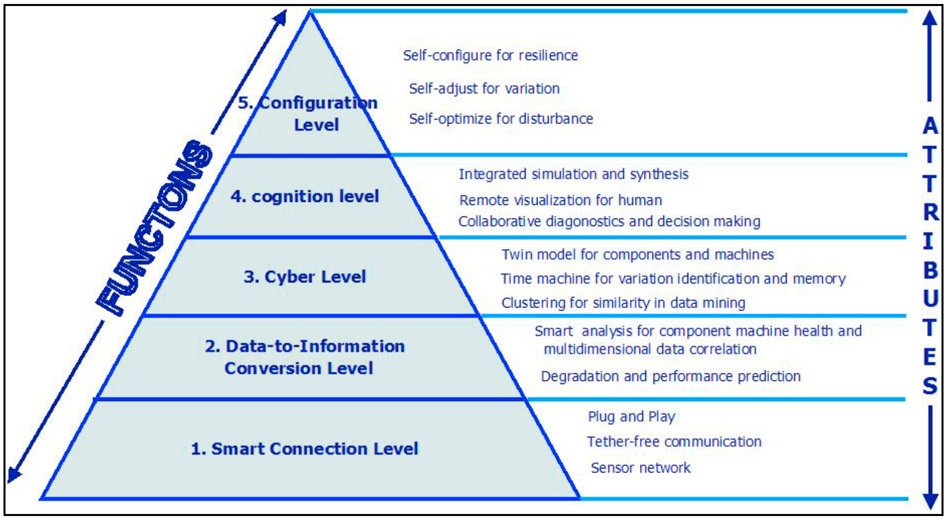

FUNCTONS
5. Configuration Level
4. cognition level
3. Cyber Level
2. Data-to-Information Conversion Level
1. Smart Connection Level
Self-configure for resilience
Self-adjust for variation
Self-optimize for disturbance
Integrated simulation and synthesis
Remote visualization for human
Collaborative diagonostics and decision making
Twin model for components and machines
Time machine for variation identification and memory
Clustering for similarity in data mining
Smart analysis for component machine health and multidimensional data correlation
Degradation and performance prediction
Plug and Play
Tether-free communication
Sensor network
ATTRIBUTES

Fig.3 5C architecture

Case Study: The manufacturing industry

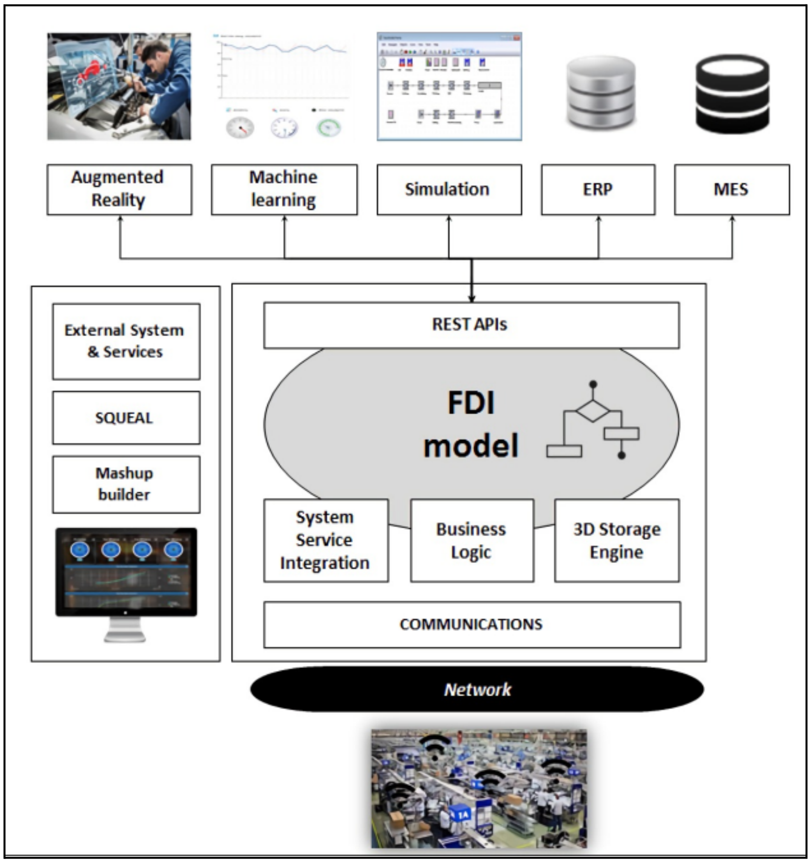


Fig.4 CPS Architecture for Manufacturing Industry

Thingworx, an application platform that bridges the physical and cyber worlds and big data management, AI-based data analytics, and cyber security, supports various industrial protocols. It also offers a dashboard that displays machine status using the mashup builder and composer layer. Factory Design and Improvement (FDI) is an all-inclusive guide on factory design and operations. Thingsworx's 5c architecture lets you connect, build, analyze, experience, and collaborate. The commercial IoT platform improves the productivity of cloud-computing-based CPS development by allowing for a simple interface with existing systems and engineering tools.

Attacks on IoT-CPS

Recent trends prove that IoT-CPS systems are prone to cyber-attacks, and their rate increases day by day. The following are some cyber-attacks.

- **Denial-of-Service (DOS):** the primary purpose of this attack is to make resources unavailable for intended use by interrupting the system forever or momentarily by flooding the target. In Distributed Denial-of-Service attacks, marks are carried out by flooding the target with unnecessary request packets from several sources.

- **Permanent Denial-of-Service (PDOS):** An attacker tries to brick an IoT device or damage its firmware (BIOS), leaving the device or system unusable. One of the PDOS attacks is USB Killer, which sucks power from the USB port until no power can be extracted from the device, providing the opportunity to replace the device.

- **Brickerbot:** malware that targets Linux IoT devices with an open telnet port and the BusyBox toolbox installed. It tries brute force using a list of known usernames and passwords, then corrupts the storage, breaks the connection, and deletes all files after acquiring access to the device.

- **Mirai Botnet:** A botnet is a network of infected devices linked together. It does DDoS using collective power (acting as a force multiplier).

- **Deauthentication Attack:** A DoS attack's primary purpose is to deauthenticate IoT-device access to the access point, causing the device to be disconnected from the network.

- **Sybil Attack:** An attacker leverages stolen or fabricated IoT device identities to compromise data integrity by dropping or forging packets in an environment where all IoT devices cannot be individually identified.

- **Man-in-the-Middle Attack:** The attacker inserts himself into communication between a user and an application to eavesdrop and appear to be a legitimate flow of information or to steal credentials and data. They are accomplished mainly by email hijacking, which involves establishing a secure connection over an unsecured channel using SSL and TSL.

- **UART (Universal Asynchronous Receiver/Transmitter) Access Attack:** A UART is a hardware device (circuit on a PCB) utilized in an asynchronous serial communication protocol used in the debugging shell or console of the device's embedded OS. An attacker modifies IoT devices using UART to evade authentication through a serial connection.

- **Worm Attack:** Worms, like viruses, multiply themselves through recursive techniques without the help of host applications and cause damage to the system. The majority of tasks are inserted through open ports.

- **Stuxnet:** a kind of malware that exploits previously undisclosed zero-day vulnerabilities in Windows. It inspects Siemen PCLs and modifies their programming, resulting in a damaging delicate component process.

- **The Black Hole Attack:** also known as the Sinkhole Attack, involves the creation of a bogus malicious node that publishes false information such as the cheapest path to attract additional nodes. Consequently, packets are sent to the malicious node, which drops them instead of routing them to the destination.

- **Energy Drain Attack:** aimed at restricted battery-powered devices, allowing these devices to drain the battery in computationally intensive activities. For example, delivering forged encrypted packets full of trash to the target node results in wasteful signature and integrity checks.

Several solutions are offered to combat the aforementioned never-ending attacks. A system administrator should regularly follow recommended procedures such as monitoring and tracking traffic and data recovery to maintain physical layer protection, including sensors, actuators, and RF identification devices. A firewall, encryption, key management, and security protocol must be implemented for transport layer security. The administrator should update control policies regularly for application-layer security, add an anomaly detection and authentication module, utilize a secure payment protocol, and use secure HTTP.

Sybil Attack

A malicious node is a single physical device with several identities, also known as a fake node-Sybil node. These assaults can be launched in three different methods, explained more below.

Taxonomy of Sybil Attack

1. **Direct or indirect communication:** Sybil node connects with legitimate node directly or through another malicious node, redirecting it to the selected Sybil node and vice versa.

2. **Identity theft or forgery:** Sybil node either steals the identity of a valid node or creates a new identity. These can be performed when the impersonated node is idle or removed from the network.

3. **Simultaneous or non-simultaneous:** Either one or multiple Sybil nodes are active simultaneously. A device can only have one functional identity. As a result, it can swap between compromised identities over time.

Influence of Sybil Attack

1. **Routing:** Sybil attack can harm multipath routing methods if the path contains malicious nodes. Because a Sybil node might appear in several locations instead of just one, it impacts the geographical routing protocol.

2. **Data Aggregation:** Instead of individual readings, networks rely on aggregate readings. A few malicious nodes cannot influence aggregate reading, whereas a Sybil node can contribute several times and hence influence the aggregate reading.

3. **Voting:** As indicated in data aggregation, a Sybil node might often contribute, influencing the outcome.

4. **Misbehavior Detection:** An administrator can identify misbehaving nodes, although it is challenging since Sybil nodes might have several identities. If one of the attacker's identities is detected as felonious, the attacker has the option of using a different identity.

5. **Fair Resource Allocation:** In a network, bandwidth is shared across devices for a specific time. By changing its identity, the attacker can disrupt the fair allocation of resources by allocating resources to the same node.

According to the study, Sybil Attack impacts IoT-CPS time synchronization at the sensor and actuator levels.

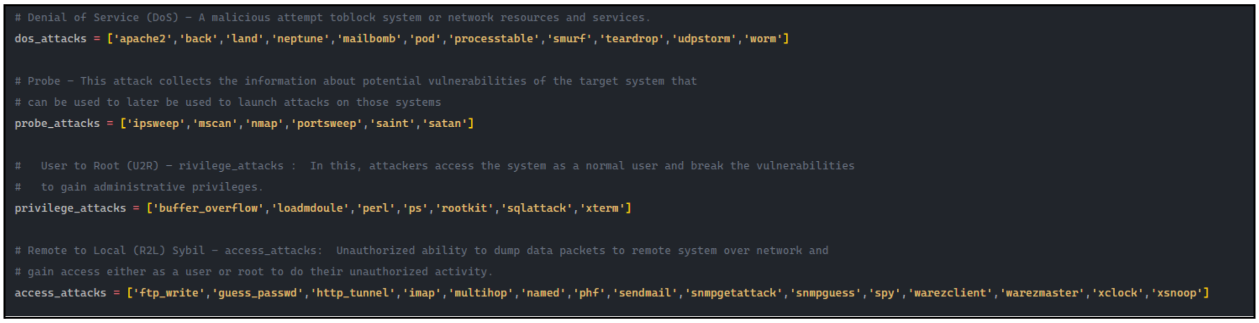
Countermeasures against Sybil attacks :

Several approaches are proposed to tackle the Sybil attack. They are classified as Prevention, Detection, and MItigate. One of the approaches is Radio resource testing (RTT) which lowers the probability of being attacked. One of the approaches to securely delivering the packets is to select the next-hop according to the trust value and energy criteria. Researchers came up with asymmetric key concepts with encryption and decryption and trusted certification to authenticate nodes, but it has a significant flaw as scalability. Some papers were published to detect Sybil nodes using RSSI (Receiver Signal Strength Indicator) and TDOA (Time Difference of arrival) to locate the Sybil node in the network. Still, it has one drawback: radio signals are susceptible to interference, and there is signal attenuation due to surroundings. Various intrusion detection systems are proposed that use swarm intelligence and ant colony optimization. Suppose we see the distribution of papers submitted on methodology to tackle Sybil attack from 2010 to 2020: RSSI (29%), Encryption (29%), AI (14%), Trust (14%), Hybrid (7%), Multikerenel (4%), Rule-based(3%). Even though there is a scope for improvement, new approaches like Blockchain and Software Defined Networks (SDN) can be used.

Experiment

Dataset:

NSL-KDD is a dataset that modified the KDD '99 dataset, which had significant flaws. The even newer version is not the perfect representation of the real-world network. But it is used as a benchmark dataset for intrusion detection methods. The target column is the type of attack, which is classified as below.



```
# Denial of Service (DoS) - A malicious attempt toblock system or network resources and services.
dos_attacks = ['apache2','back','land','neptune','mailbomb','pod','processtable','smurf','teardrop','udpstorm','worm']

# Probe - This attack collects the information about potential vulnerabilities of the target system that
# can be used to later be used to launch attacks on those systems
probe_attacks = ['ipsweep','mscan','nmap','portsweep','saint','satan']

#   User to Root (U2R) - rivilege_attacks :  In this, attackers access the system as a normal user and break the vulnerabilities
#   to gain administrative privileges.
privilege_attacks = ['buffer_overflow','loadmdoule','perl','ps','rootkit','sqlattack','xterm']

# Remote to Local (R2L) Sybil - access_attacks:  Unauthorized ability to dump data packets to remote system over network and
# gain access either as a user or root to do their unauthorized activity.
access_attacks = ['ftp_write','guess_passwd','http_tunnel','imap','multihop','named','phf','sendmail','snmpgetattack','snmpguess','spy','warezclient','warezmaster','xclock','xsnoop']
```

Fig.5 NSL KDD attack types categorization

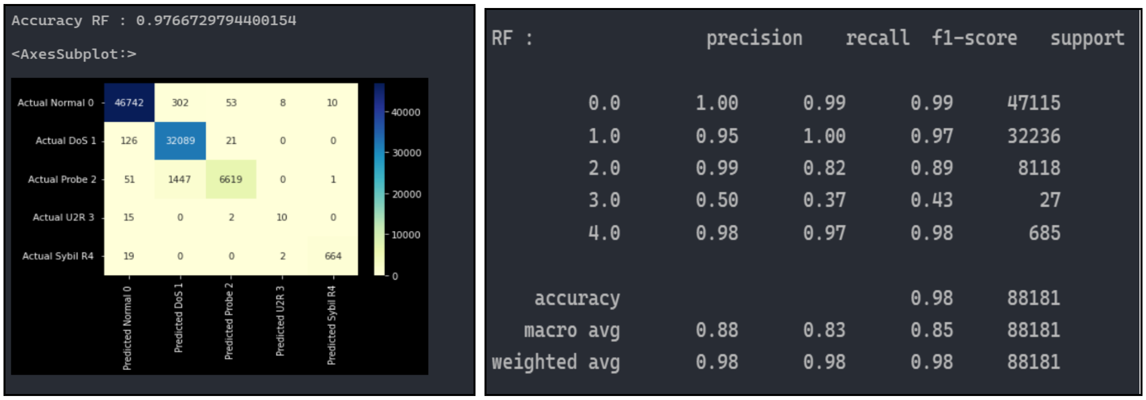
Languages/Tools/Packages used:

- Python
- Google Colab, Jupyter Notebooks
- Pandas, Numpy, MatplotLib, Seaborn, Tensorflow, Keras, Scikit-Learn

Result:

1. Random Forest Algorithm:

   Random Forest is a supervised machine learning model which uses bagging and boosting concepts to predict the class. The majority output of decision trees can give the verdict.



```
Accuracy RF : 0.9766729794400154

<AxesSubplot:>
```

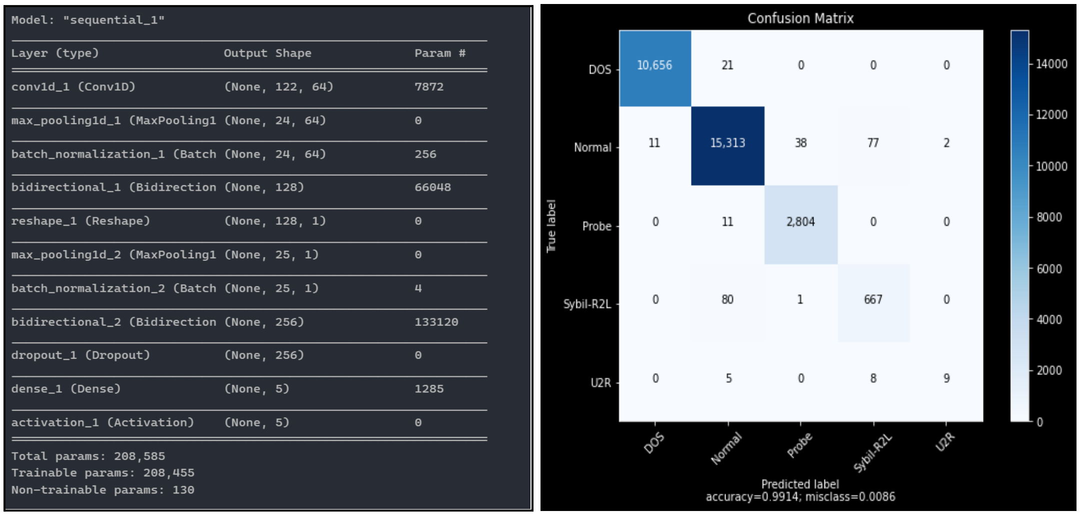
	Predicted Normal 0	Predicted DoS 1	Predicted Probe 2	Predicted U2R 3	Predicted Sybil R4
Actual Normal 0	46742	302	53	8	10
Actual DoS 1	126	32089	21	0	0
Actual Probe 2	51	1447	6619	0	1
Actual U2R 3	15	0	2	10	0
Actual Sybil R4	19	0	0	2	664

```
RF :            precision    recall  f1-score   support

         0.0       1.00      0.99      0.99     47115
         1.0       0.95      1.00      0.97     32236
         2.0       0.99      0.82      0.89      8118
         3.0       0.50      0.37      0.43        27
         4.0       0.98      0.97      0.98       685

    accuracy                           0.98     88181
   macro avg       0.88      0.83      0.85     88181
weighted avg       0.98      0.98      0.98     88181
```

Fig.6 Random Forest: confusion matrix & classification report

2. CNN Bi-LSTM model:

J.Sinha et al. [11] presented the paper in which they built the deep learning model, which is a blend of CNN (Convolution Neural Network) and Bi-LSTM (Bi-Directional Long Short Term Memory)



```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv1d_1 (Conv1D)            (None, 122, 64)           7872
_________________________________________________________________
max_pooling1d_1 (MaxPooling1 (None, 24, 64)            0
_________________________________________________________________
batch_normalization_1 (Batch (None, 24, 64)            256
_________________________________________________________________
bidirectional_1 (Bidirection (None, 128)               66048
_________________________________________________________________
reshape_1 (Reshape)          (None, 128, 1)            0
_________________________________________________________________
max_pooling1d_2 (MaxPooling1 (None, 25, 1)             0
_________________________________________________________________
batch_normalization_2 (Batch (None, 25, 1)             4
_________________________________________________________________
bidirectional_2 (Bidirection (None, 256)               133120
_________________________________________________________________
dropout_1 (Dropout)          (None, 256)               0
_________________________________________________________________
dense_1 (Dense)              (None, 5)                 1285
_________________________________________________________________
activation_1 (Activation)    (None, 5)                 0
=================================================================
Total params: 208,585
Trainable params: 208,455
Non-trainable params: 130
```

Confusion Matrix

True label	DOS	Normal	Probe	Sybil-R2L	U2R
DOS	10,656	21	0	0	0
Normal	11	15,313	38	77	2
Probe	0	11	2,804	0	0
Sybil-R2L	0	80	1	667	0
U2R	0	5	0	8	9

Fig.7 CNN BiLSTM: structure & confusion matrix

Conclusion

During my study, I discovered that crucial distribution, RSSI, location, and trust-based methodology presented intrusion detection and prevention approaches for the Sybil Attack. But, these solutions increase the workload on the system, which affects their energy consumption. The Deep Neural Network is the best choice among any intrusion detection, with higher accuracy and detection rate. The lack of real-world datasets is a key disadvantage of any intrusion detection system. Future studies include the extension of devices that can monitor and analyze the network. The generated data can be fed to construct the intrusion detection model, which will help predict the attack in advance, prevent changes in the current state of the network & dump the unauthenticated malicious data packets, which have to happen in real-time.

Bibliography

[1] C. Greer, M. Burns, D. Wollman, and E. Griffor, “Cyber-physical systems and internet of things,” National Institute of Standards and Technology, Mar. 2019 [Online].
Available: http://dx.doi.org/10.6028/NIST.SP.1900-202

[2] S. Choi, G. Kang, C. Jun, J. Y. Lee, and S. Han, “Cyber-physical systems: a case study of development for manufacturing industry,” International Journal of Computer Applications in Technology, vol. 55, no. 4. Inderscience Publishers, p. 289, 2017 [Online].
Available: http://dx.doi.org/10.1504/IJCAT.2017.086018

[3] V. Ponnusamy, N. D. Regunathan, P. Kumar, R. Annur, and K. Rafique, “A Review of Attacks and Countermeasures in Internet of Things and Cyber Physical Systems,” Industrial Internet of Things and Cyber-Physical Systems. IGI Global, pp. 1–24, 2020 [Online].
Available: http://dx.doi.org/10.4018/978-1-7998-2803-7.ch001

[4] A. Arshad, Z. Mohd Hanapi, S. Subramaniam, and R. Latip, “A survey of Sybil attack countermeasures in IoT-based wireless sensor networks,” PeerJ Computer Science, vol. 7. PeerJ, p. e673, Sep. 22, 2021 [Online].
Available: http://dx.doi.org/10.7717/peerj-cs.67

[5] J. Newsome, E. Shi, D. Song, en A. Perrig, “The Sybil attack in sensor networks: analysis & defenses,” Third International Symposium on Information Processing in Sensor Networks, 2004. IPSN 2004, bll 259–268, 2004.[Online].
Available: https://bit.ly/34Vwsz2

[6] J. Douceur, “The Sybil Attack,” in Proceedings of 1st International Workshop on Peer-to-Peer Systems (IPTPS), 2002.[Online].
Available: https://bit.ly/3L2HTVg

[7] A. Mehbodniya, J. L. Webber, M. Shabaz, H. Mohafez, and K. Yadav, “Machine Learning Technique to Detect Sybil Attack on IoT Based Sensor Network,” IETE Journal of Research. Informa UK Limited, pp. 1–9, Dec. 08, 2021 [Online].
Available: http://dx.doi.org/10.1080/03772063.2021.2000509

[8] K. Debasis, M. P. Singh, P. Kumar, and S. Bhaskar, “Detection of Sybil Nodes in Wireless Sensor Networks,” Indian Journal of Science and Technology, vol. 10, no. 3. Indian Society for Education and Environment, Jan. 23, 2017 [Online].
Available: http://dx.doi.org/10.17485/ijst/2017/v10i3/110641

[9] G. Mohi-ud-din, “NSL-KDD.” IEEE DataPort, Dec. 29, 2018 [Online].
Available: https://ieee-dataport.org/documents/nsl-kdd

[10] D. Kumari, K. Singh, and M. Manjul, “Performance Evaluation of Sybil Attack in Cyber-Physical System,” Procedia Computer Science, vol. 167. Elsevier BV, pp. 1013–1027, 2020 [Online].
Available: http://dx.doi.org/10.1016/j.procs.2020.03.401

[11] J. Sinha & M. Manollas, ‘Efficient deep CNN-BiLSTM model for network intrusion detection, And Proceedings of the 2020 3rd International Conference on Artificial Intelligence and Pattern Recognition, 2020, pp. 223–231,2020 [Online]
Available: http://dx.doi.org/10.1145/3430199.3430224

.