Deep Learning Techniques for Security in Edge Computing: A Detailed Survey

R.Anusuya[1], Dr.D. Karthika Renuka [2], S. Bhuvaneshwari[3]
{anusuya12@gmail.com [1], dkr.it@psgtech.ac.in[2] bhuvana1096@gmail.com[3]}

Department of Information Technology,PSG College of Technology,Coimbatore[1,2,3]

Abstract. Massive amounts of data are generated instantly and as computing power gets increased subsequently the performance of cloud computing is dissatisfying. The security and privacy concerns of the user is also a serious issue. Edge computing (EC) is taken into account in recent years to resolve these issues. The major goal of this study is to know how well edge computing corresponds to the cloud and notably improves the overall performance. In the context of edge computing, the paper also shows how effective deep learning approaches are for security.

Keywords: Edge Computing, Deep Learning, Autoencoders..

1 Introduction

While demand for Internet of Things (IoT) is on the rise, we are entering the post cloud era where data of large volumes are produced and many devices are employed to process the information at the edge. A prediction by Cisco System estimates that a 5.3 billion global population will connect to the Internet by 2023. As some applications require data privacy and short response time, cloud computing is not feasible. EC becomes the saviour for such applications.

Sensors, wearables, mobile devices, and other IoT technologies have revolutionised how we live and work in recent years. At present, there are more than 10 billion active IoT devices. It is forecasted that the number will shoot up beyond 25 billion in 2030. As per the McKinsey Global Institute, the IoT might generate $11 trillion in annual economic value by 2025. Connected gadgets serve as the foundation for smart systems like those for smartwatches, smart cities, automobiles, appliances, and smart metering. People have become more reliant on digital along with smart gadgets every day, as well as new connected devices are constantly appearing, although only a tiny portion of IoT data is now being used.

Why do we need edge computing?

EC offered a solution to these problems by physically shifting processing out from the central server and closer to the network and data source edges. Thus, processing at the edge would yield reliable, efficient, and shorter response time. Partially performing the calculation either upon the device or on a node near the data provider may reduce the volume that needs to be transferred to the cloud lowering latencies and improving response time. EC could potentially be used to extract and minimize the features in a dataset.

ICCAP 2021, December 07-08, Chennai, India

DOI 10.4108/eai.7-12-2021.2314539

	EDGE COMPUTING	CLOUD COMPUTING
Reliability	High since each edge server can be terminated separated	Low since it's a centralized module
Data privacy	Secure	May leak on the internet
Network traffic	Small as only warning message will be transmitted	Large as all the images and video are transmitted
Network latency	Less than ten milliseconds	Between 50 to 500 milliseconds

Table 1. Comparison between edge, cloud models

Why do we need Deep Learning?

Deep Learning (DL) has experimentally found promising outcomes in every range of fields, particularly while enormous amounts of data seem to be accessible. It does have a tremendous opportunity in the IoT since it can learn important features by converting data into hierarchical abstract representations. Autoencoders (AE) is an unsupervised learning algorithm that could be used to understand how to encode data effectively.

The following is how the remaining part of the paper is structured:Section II discusses the security aspects and Section III provides background details and Section IV refers to the related work. The application of autoencoders is mentioned in Section V. Section VI brings the paper to a conclusion.

Table II. List of Acronyms

IoT	Internet of Things
EC	Edge Computing
DL	Deep Learning
AE	Autoencoder
ML	Machine Learning
FL	Federated Learning
CNN	Convolutional Neural Networks
DoS	Denial of Service
DDoS	Distributed Denial of Service
RFID	Radio-Frequency Identification
MAC	Medium Access Control
MITM	Man-in-the-middle Attack
HCP	Dynamic Host Configuration Protocol
ARP	Address Resolution Protocol
DNS	Domain Name System

2. Security aspects

As massive data are involved in real-time for generation and processing. The handling of these data will be a major security concern in recent years. As we all know that images are one important type of data available on every device. In order to make it meaningful information, we have to process the images with keeping privacy in our mind. As the images contain private information, end-users may have a problem while uploading the image as they are more concerned about privacy in today's world. Moreover, it is a matter of issue regarding eavesdropping, interceptions, or other unauthorized access.

1) Possible Attacks

The significant security dangers and difficulties influencing the EC framework are as follows:

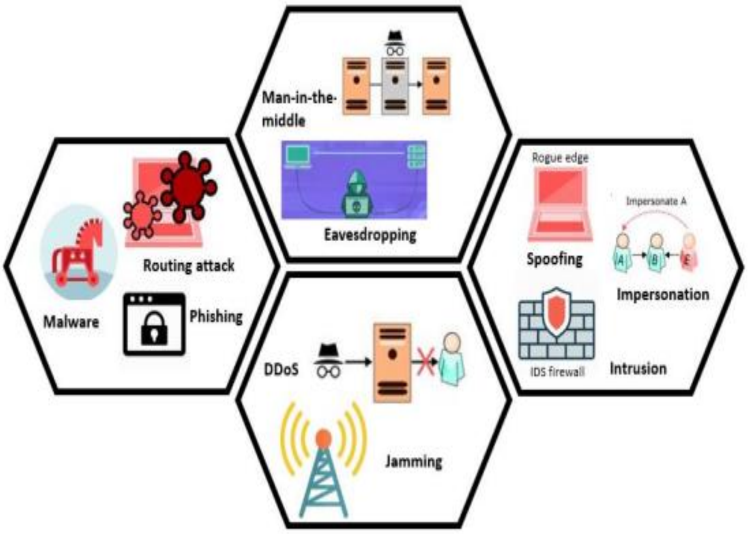

Fig 1. Attacks in EC framework

i. Distributed Denial of Service (DDoS)

In Denial of Service (DoS), strategies towards obstructing or purposefully corrupting both accessibility as well as association with the system or its resources is been executed. Access is been forestalled for legitimate devices when the server of the resource is affected. As in the case of DDoS, the attacker is capable to altogether expand the adequacy of the attack.

Because they have fewer computing resources for security, deploy heterogeneous firmware, and the majority of nodes are not mutually authorized, edge servers are more vulnerable to DDoS attacks than cloud servers.

ii. Eavesdropping

The act of intercepting a transmission packet sent through a network is known as eavesdropping, sometimes known as sniffing or snooping. This type of assault is possible with radio-frequency identification (RFID) devices. The confidentiality of systems deteriorates as an attacker acquires access to critical information such as node identification and configuration.

iii. Malware Injection

This attack infects the system with a malicious script or software. The intruder persuades the EC to treat the new service deployment as a separate instance of the system. A legitimate user request is routed to this service, and the attacker's code is subsequently executed. Malware such as Trojans, worms, and viruses cause data leakage, battery drain, and performance degradation.

iv. Jamming

Malicious users utilize bogus signals to disrupt ongoing communication on the Edge network. The attacker interferes with wireless sensor node radio frequencies, jamming or delaying transmission. During failed communication attempts, this reduces the bandwidth and storage capacity of the nodes.

v. Spoofing

In this case, an impersonating node poses as a valid one throughout the Edge network with a forged identification, such as a MAC address or an RFID tag. Once within the system, DDoS and Man-in-the-middle Attack (MITM) assaults are frequently assisted.

vi. Man-in-the-middle Attack

Whenever a group of devices within the system exchange keys, a secure communication channel is established. A MITM attack occurs when an attacker intercepts communications and exchanges keys with some devices individually. The attacker gets hold of the transmission medium between authorized network devices and sends out buzzing signals to track, eavesdrop on, or modify communication among the network's Edge Devices. Spoofing of the Dynamic Host Configuration Protocol (DHCP), cache poisoning of the Address Resolution Protocol (ARP), Domain Name System (DNS) spoofing, Port Stealing, and Session Hijacking are all instances of this form of attack.

vii. Routing

The attack's goal is to sabotage the routing systems. A Sinkhole attack uses the routing metric to divert traffic from a particular region to the compromised node, trying to make it look authentic to other nodes.

viii. Phising

Phishing entails impersonating a respectable or trustworthy source in order to obtain confidential piece of information from the victim. Typosquatting, which includes utilizing frequent or undetectable errors in spelling, and cybersquatting, that involves using already available domain names for fraudulent purposes, make it simpler to entice or mislead users.

2) Securing EC

As EC networks are frequently resource constrained, ML or cryptographic security mechanisms are challenging to implement. To assist EC systems in making intelligent and informed judgments for recognizing and responding to assaults, ML and DL techniques are used. After being trained on suitable datasets, these techniques are implemented on edge servers or devices.

They enable the EC environment to intelligently gain knowledge on attack vectors and behaviors with no need for explicit design. The ML method chosen for DDoS attack detection would be influenced by the quantity of dataset, the number of attributes, duration for training, precision, and representativeness of the output. Memory and processing resource limits are two important aspects of the EC environment. Also consider if the algorithm is being trained upon real-time data as well as whether there is labeled or unlabeled data, among other parameters.

3. Background Details

*A. **Deep Learning***

DL is a subset of the machine learning (ML) method that employs a series of layers, each of which performs a non-linear transformation. Because of its ability to learn complex models and perform representation learning, DL has received significant attention in recent years. Learning is done through data representation instead of directly performing on data. An abstract representation is obtained from the data which is useful to learn useful features and subsequentially used for ML. Convolutional neural networks (CNN), AE, and recurrent neural networks (RNN) are few examples of DL architectures.

*B. **Autoencoder***

A feedforward neural network with the same input and output is known as an AE. Before reconstructing the output from the representation, the input is compressed into a lower-dimensional code. The latent-space representation, commonly known as the code, is a concise "summary" or "compression" of the input.

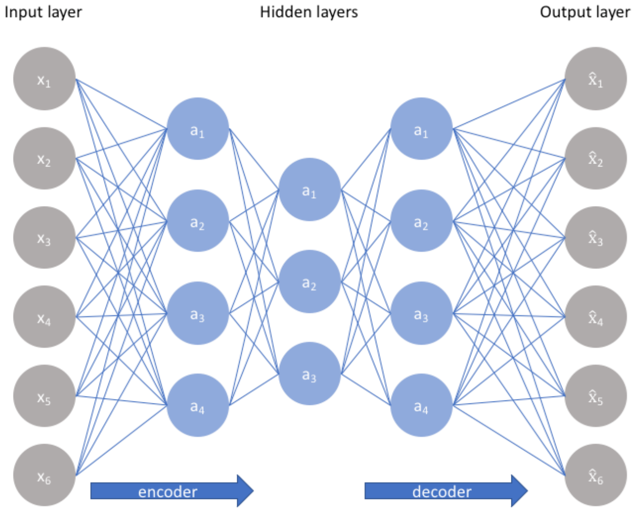

Fig 2. Visual representation of AE

An AE is made up of three parts namely an encoder, a code, and a decoder. The encoder compresses the data and generates a code from it. The decoder uses only this code to reconstruct the input.

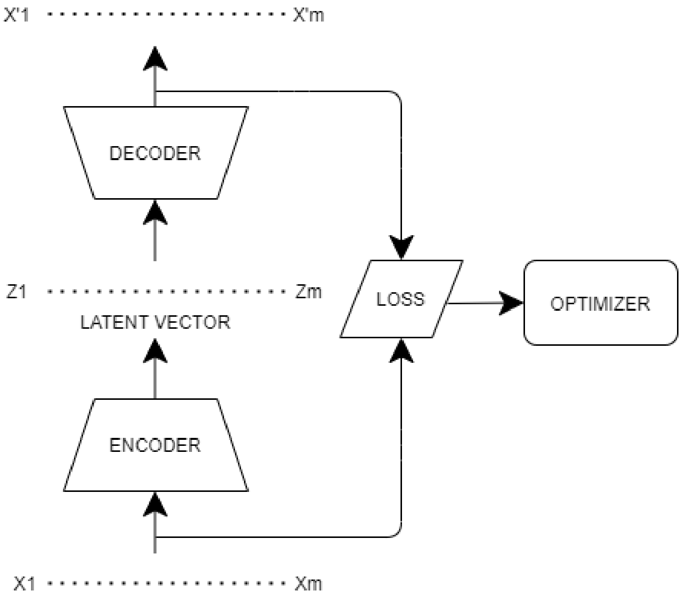
X'1 .... X'm
DECODER
Z1 .... Zm
LATENT VECTOR
ENCODER
X1 .... Xm
LOSS
OPTIMIZER

Fig 3. AE Architecture

C. *Advantages over Federated Learning (FL)*

The server is responsible for updating the gradients and training the model in FL, therefore the edge device must rely on it. Without the use of a server, each edge device can train the AE individually.

FL ensures the security of end users' data by employing differential privacy techniques or secure aggregation, both of which involve extra costs due to encryption or secret sharing. In AE, the privacy of end users' data is ensured by transmitting latent vectors without the need for further encryption.

4. Related Work

Fagbohungbe et al., [1] proposed a novel EC framework along with an AE to address end users' concerns about security. But training a robust AE will take time. Pattanayak et al., [2] Provided a Fuzzy logic membership function and an autoencoder neural network to encrypt or hide sensitive information and also to securely send the data to various organizations. Fuzzy logic is not always accurate. Alguliyev R. M. et al., [3] propounded a CNN and modified sparse denoising AE to alter the original data inorder to keep the data and knowledge private even after the mining process. To hamper the data and utilize it as input for training an AE to denoise data, basic stochastic mapping is required.

Singh, S et al., [4] studied the various ML techniques for securing privacy in EC. The author has also explained the various DL methods used for each attack. Ghosh, A. M. et al., [5] analyzed using DL for the edge-cloud platform. Kaissis, G et al., [6] summaries about privacy preserving DL on multi-institutional medical imaging. Wang, X., et al., [8] explains the DLinferences in edge and the corresponding edge security aspects. Boulemtafes, A et al., [13] summaries about different privacy-preserving DL techniques and the recent solutions.

5. Application of Auto encoders

In recent years, AE has been used for image denoising and dimensionality reduction.

A. *Image Denoising*

A noisy image is something in which the image is been corrupted or some noise is been added to it. Inorder to acquire suitable information, image denoising is performed.

B. *Dimensionality Reduction*

A significant set of attributes, or features, are frequently encountered in ML problems. These are very common in ML tasks where the data set is large. On the other hand, it becomes more difficult, as the dimensionality grows larger. Working with these data set is difficult and time consuming. As a result, the ML algorithm's effectiveness decreases. Hence "dimensionality reduction” denotes a reduction in the number of features or attributes that are taken into account.

C. *Feature Extraction*

The encoding component of AE aids in the learning of important hidden features present in input data, thereby reducing reconstruction error. A new set of combinations of original features is generated during encoding.

*D. **Privacy***

To enhance the security of data, AE encrypts the raw data before transforming it to a latent vector. To keep security issues in mind, raw data is never transferred. It provides a level of security comparable to standard encryption. Even if an attacker acquires the edge device, deducing the decoder component from that of the encoder portion on the edge device is exceedingly difficult.

6. Conclusion

This paper summarizes the potentials of EC along with DL. EC came to the rescue to process the data over the networks' edge near where it is collected. EC not only decreases the computational and connectivity burden on IoT networks and cloud servers, but also allows for privacy preservation by transmitting only the information that is necessary rather than raw data. The concept of DL specifically AE is analysed in the paper. To reduce the amount of features and data size, the AE's encoder is installed at the edge. Similarly, the AE's decoder can use the reduced data directly for ML tasks like classification, or to restore the original data. Anticipated that this survey will spark more discussion and research into the integration of DL with Edge, which will help to advance future applications and services.

References

[1] Fagbohungbe, O., Reza, S. R., Dong, X., & Qian, L. (2020). Efficient Privacy Preserving Edge Computing Framework for Image Classification. *arXiv preprint arXiv:2005.04563*.

[2] tanayak, S., & Ludwig, S. A. (2019, June). Improving Data Privacy Using Fuzzy Logic and Autoencoder Neural Network. In *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1-6). IEEE.

[3] Alguliyev, R. M., Aliguliyev, R. M., & Abdullayeva, F. J. (2019). Privacy-preserving deep learning algorithm for big personal data analysis. *Journal of Industrial Information Integration*, *15*, 1-14.

[4] Singh, S., Sulthana, R., Shewale, T., Chamola, V., Benslimane, A., & Sikdar, B. (2021). Machine learning assisted security and privacy provisioning for edge computing: A survey. *IEEE Internet of Things Journal*.

[5] Ghosh, A. M., & Grolinger, K. (2019, May). Deep learning: Edge-cloud data analytics for iot. In *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)* (pp. 1-7). IEEE.

[6] Kaissis, G., Ziller, A., Passerat-Palmbach, J., Ryffel, T., Usynin, D., Trask, A., & Braren, R. (2021). End-to-end privacy preserving deep learning on multi-institutional medical imaging. *Nature Machine Intelligence*, *3*(6), 473-484.

[7] Lee, S. J., Yoo, P. D., Asyhari, A. T., Jhi, Y., Chermak, L., Yeun, C. Y., & Taha, K. (2020). IMPACT: Impersonation attack detection via edge computing using deep autoencoder and feature abstraction. *IEEE Access*, *8*, 65520-65529.

[8] Wang, X., Han, Y., Leung, V. C., Niyato, D., Yan, X., & Chen, X. (2020). Convergence of edge computing and deep learning: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, *22*(2), 869-904.

[9] Subramaniam, P., & Kaur, M. J. (2019, March). Review of security in mobile edge computing with deep learning. In *2019 Advances in Science and Engineering Technology International Conferences (ASET)* (pp. 1-5). IEEE.

[10] Marchisio, A., Hanif, M. A., Khalid, F., Plastiras, G., Kyrkou, C., Theocharides, T., & Shafique, M. (2019, July). Deep learning for edge computing: Current trends, cross-layer optimizations, and open research challenges. In *2019 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)* (pp. 553-559). IEEE.

[11] Hagan, M., Siddiqui, F., & Sezer, S. (2019, August). Enhancing security and privacy of next-generation edge computing technologies. In *2019 17th International Conference on Privacy, Security and Trust (PST)* (pp. 1-5). IEEE.

[12] Zhang, J., Chen, B., Zhao, Y., Cheng, X., & Hu, F. (2018). Data security and privacy-preserving in edge computing paradigm: Survey and open issues. *IEEE access*, *6*, 18209-18237.

[13] Boulemtafes, A., Derhab, A., & Challal, Y. (2020). A review of privacy-preserving techniques for deep learning. *Neurocomputing*, *384*, 21-45.

[14] Cisco Annual Internet Report (2018–2023) White Paper

[15] M. James, M. Chui, P. Bisson, J. Woetzel, R. Dobbs, J. Bughin, and D. Aharon. The internet of things: Mapping the value beyond the hype. McKinsey Global Institute, 2015.