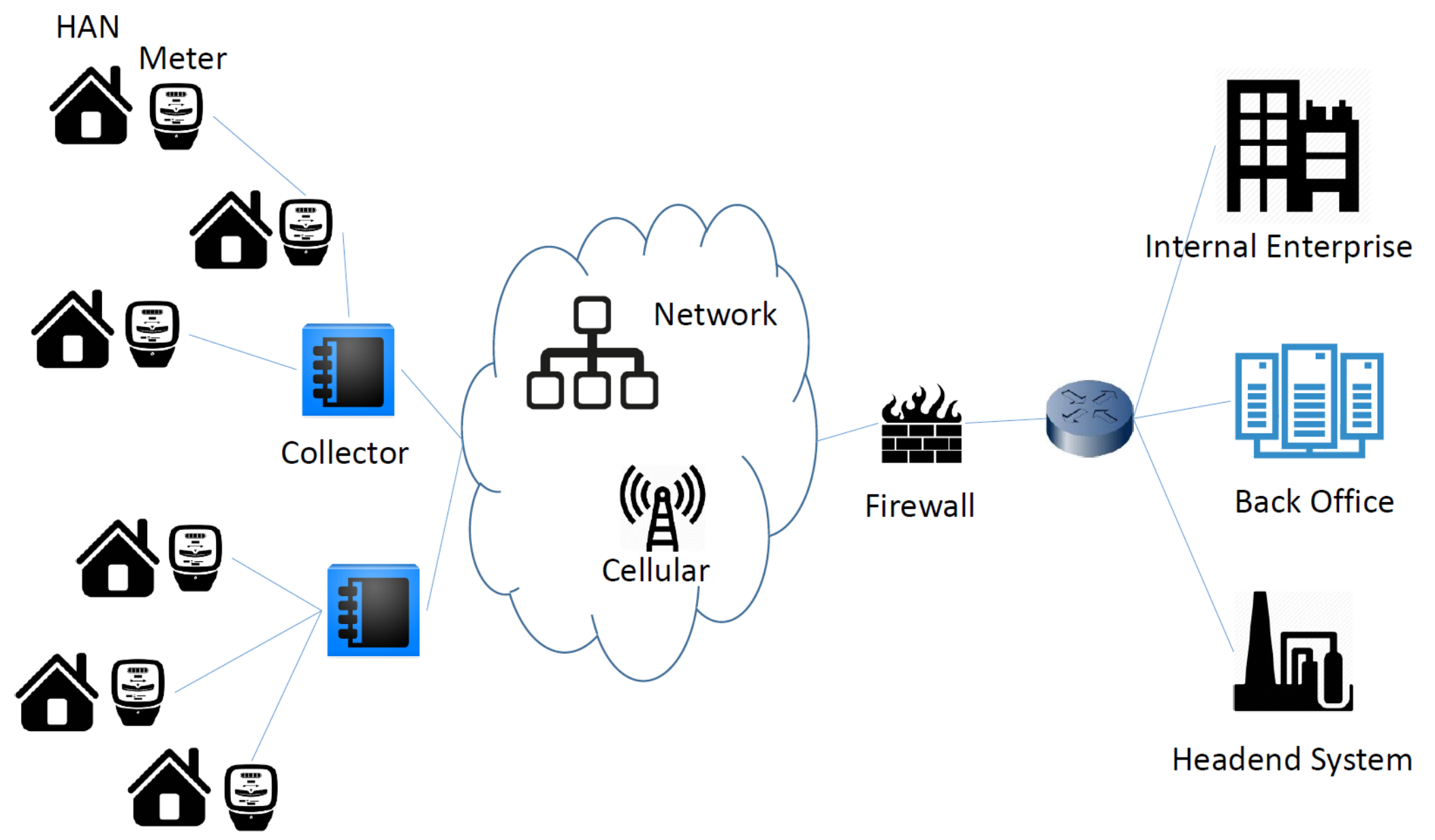
Smart Fraud Detection in Smart Metering System of AMI Networks

A H M Jakaria, Mehedi Hasan, and Douglas A Talbert, Dept. of Computer Science, Tennessee Tech University

Research Problem

- Advanced metering infrastructure (AMI) is a critical part of a modern smart grid that performs the bidirectional data flow of sensitive power information such as smart metering data and control commands.
- While smart meter data helps to improve the overall performance of the grid in terms of efficient energy management, it has also made the AMI an attractive target of cyberattackers with a goal of stealing energy.
- We propose a novel technique to detect fraudulent data from smart meters based on energy consumption patterns of the consumers by utilizing deep learning techniques.



A typical AMI infrastructure.

Challenge and Objective

- Smart meters have several vulnerabilities that are exploited by cyberattackers to manipulate the collected data [1], [2].
- One of the biggest challenges is the detection and prevention of electricity energy theft.
- We propose a machine learning based approach to address the problem, which is the first of its kind to the best of our knowledge.

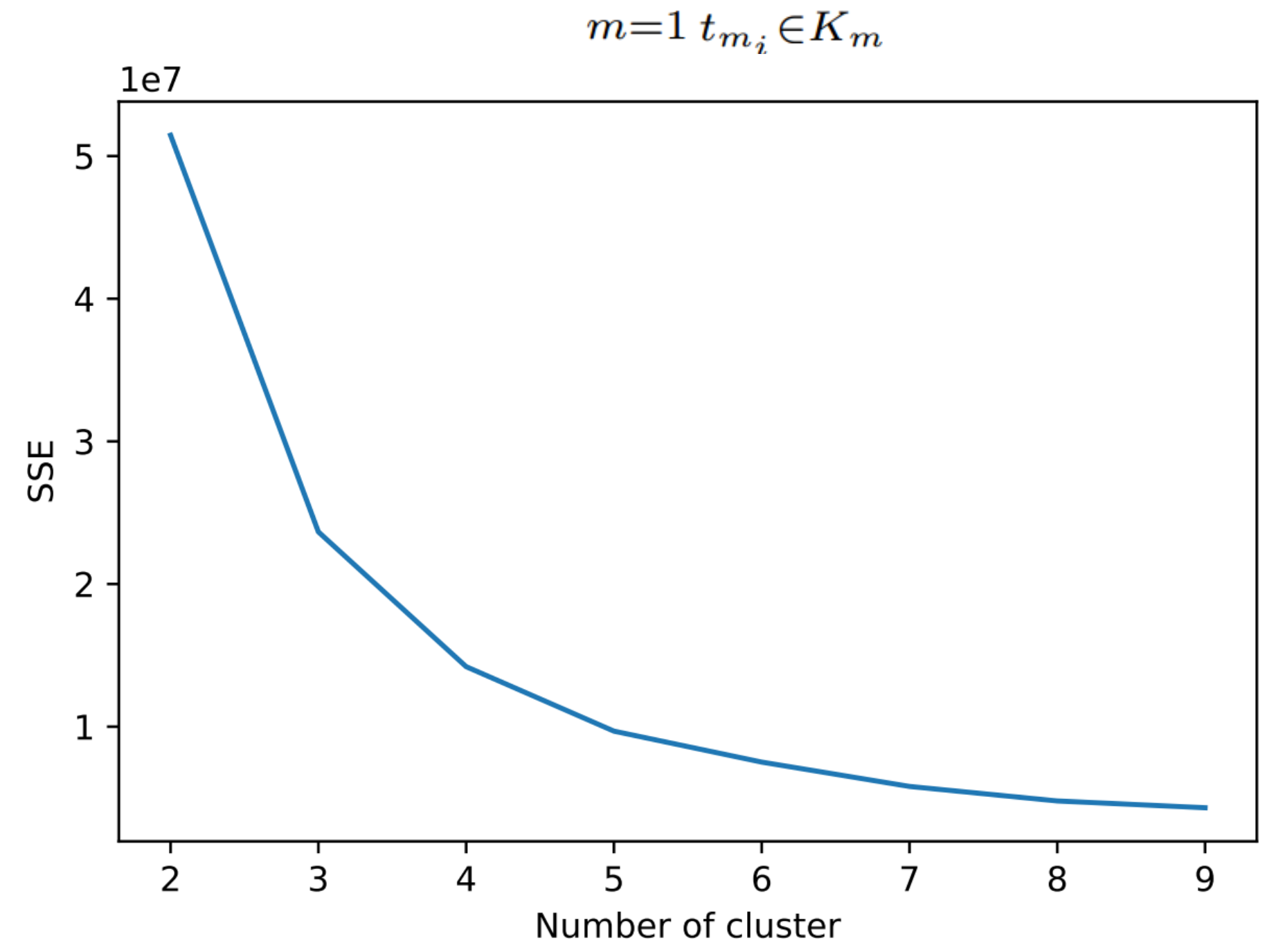
Threat Model

- Attackers in AMI: (1) Customers - Customers have been the primary adversaries. (2) Organized crime - Professional hackers exploit the extended computing and network features. (3) Utility company insiders - Dishonest or disgruntled employees in the utility companies may take part.
- Targets of Threats: (1) Smart meters - Smart meters are the most attacked components in the AMI. (2) Communications network: Usage data may be tampered after recording or during transmission. (3) Data collector: Data collectors may have remote disconnect functions, which can be exploited by attackers to create power outages [3].

Anomaly Detection: Unsupervised Technique

- The dataset contains energy consumption of users of different categories.
- We first create clusters of users with similar consumption behavior.
- We run *k*−means algorithm on our dataset, which provides us with such clusters based on time of the day and the amount of electricity consumed.
- We run k−means for different values of k, and choose the best one based on minimum sum of squared distance of the data points from corresponding centroids:

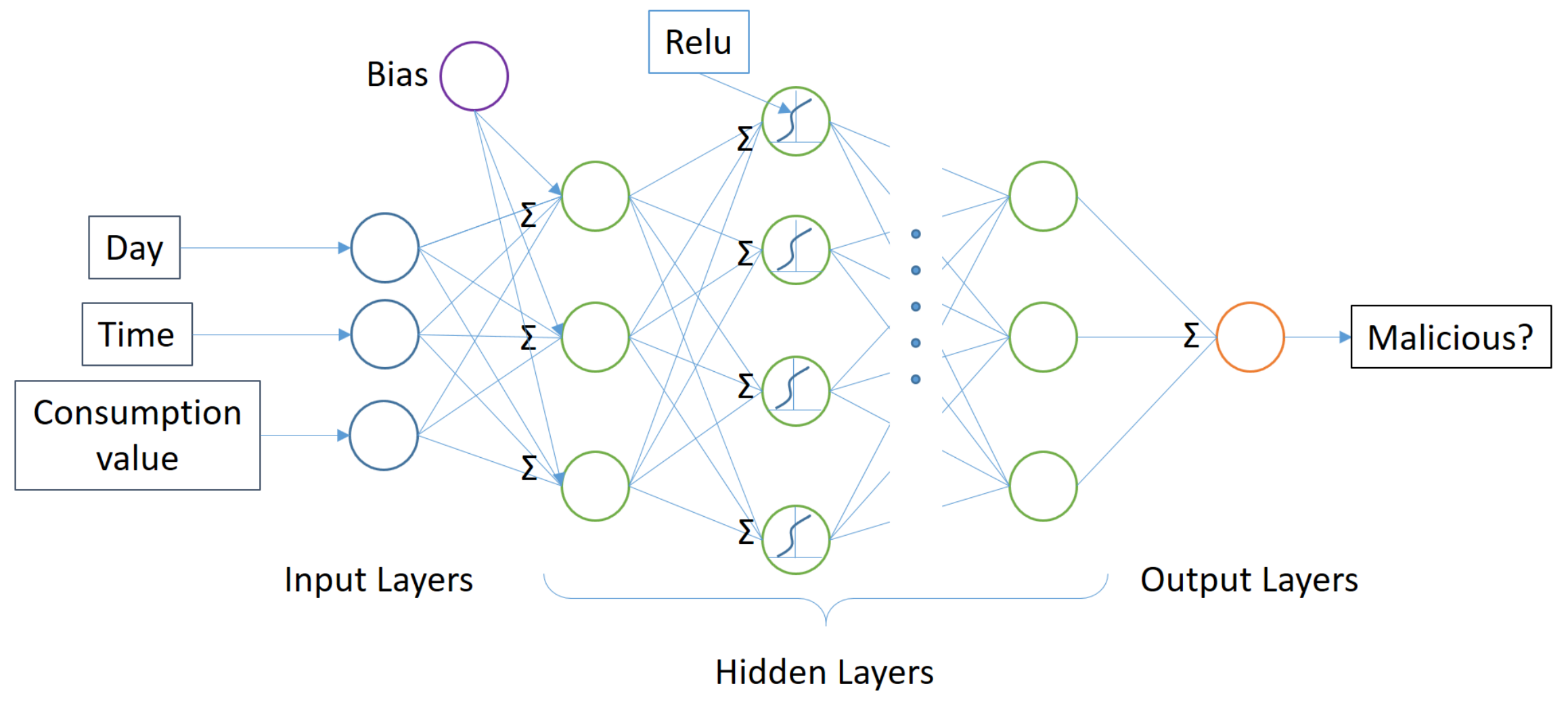
$$SumOfSqDist = \sum_{m=1}^{k} \sum_{t_{m_i} \in K_m} (C_m - t_{m_i})^2$$



We utilized the elbow method for choosing optimal number of cluster. *k* = 5 was chosen.

Anomaly Detection: Supervised Technique

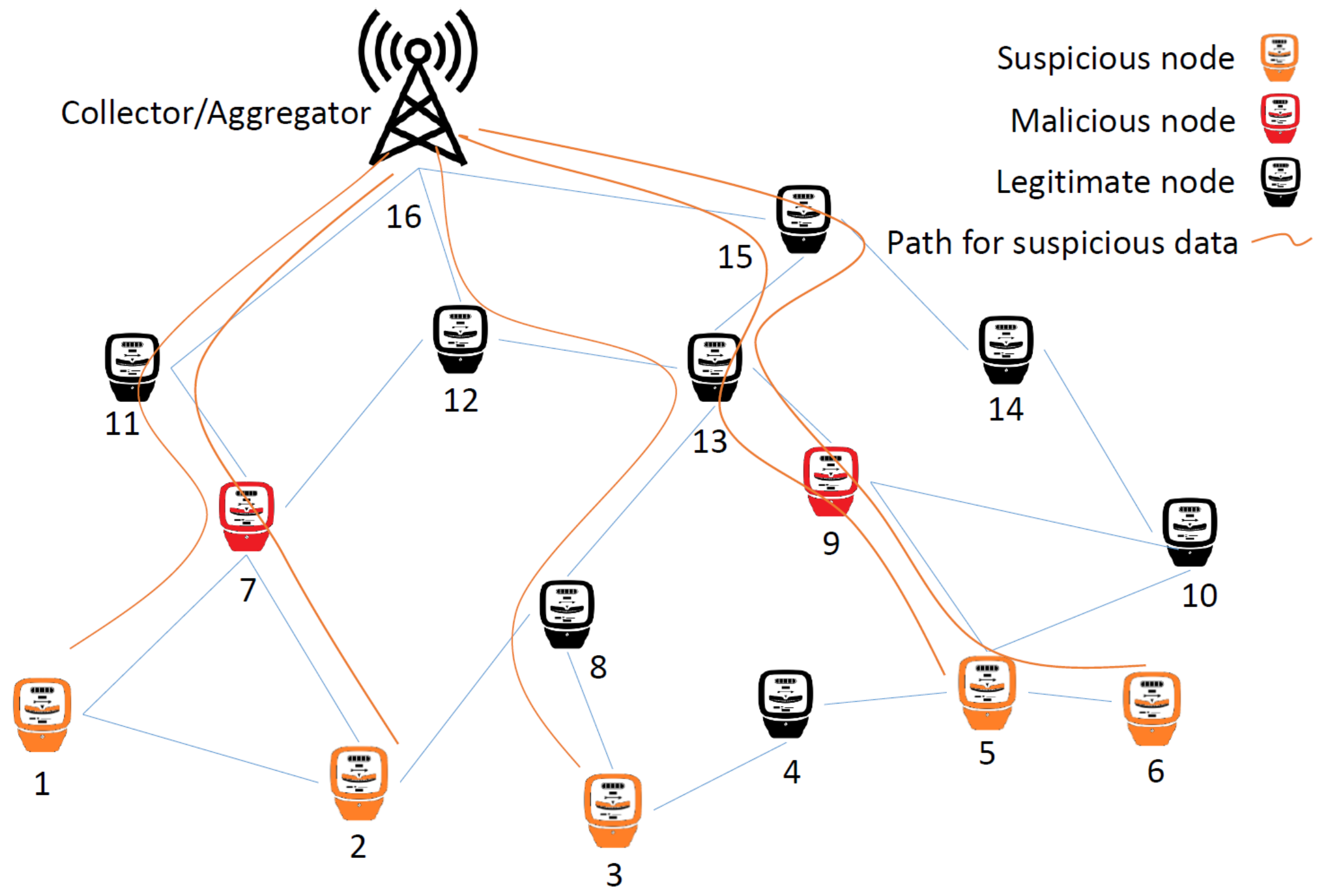
- Within each cluster, we create a dataset for training our supervised classifier.
- A 'label' attribute was introduced to the dataset, which identifies whether a data record is malicious or legitimate.



- We trained the classifier based on a multi-layer perceptron, and compared the results with several other techniques.
- We ran the model for a maximum of 200 epochs or until convergence, where in each epoch, the input samples are shuffled.
- We used the day, time, and the consumption value to learn a general pattern for the consumption.

Suspicious Node Detection

- We assume that the smart meters are connected with other meters in a mesh topology, where intermediate nodes (meters) relay the data collected by its child node to the upper level.
- If the intermediate nodes are compromised, they can be used to alter the legitimate meter data to launch an attack. Some malicious nodes may deliberately perform attacks on some other nodes.
- The compromised or malicious nodes can alter the meter data coming from other nodes. This is performed through bypassing the integrity protection schemes, if any.
- In our attack model, we consider two strategies of a malicious node in the mesh AMI network: (1) Changing any data going through itself. (2) Changing selective data from particular nodes.
- We propose two different algorithms to detect the malicious nodes in both the strategies.



A mesh network of AMI, which has some suspicious nodes detected by our machine learning model (yellow). Their data reached the aggregator through some intermediate nodes (red), which may also be malicious, as suggested by our algorithms.

Dataset: Data Collection

Data Used for Classification

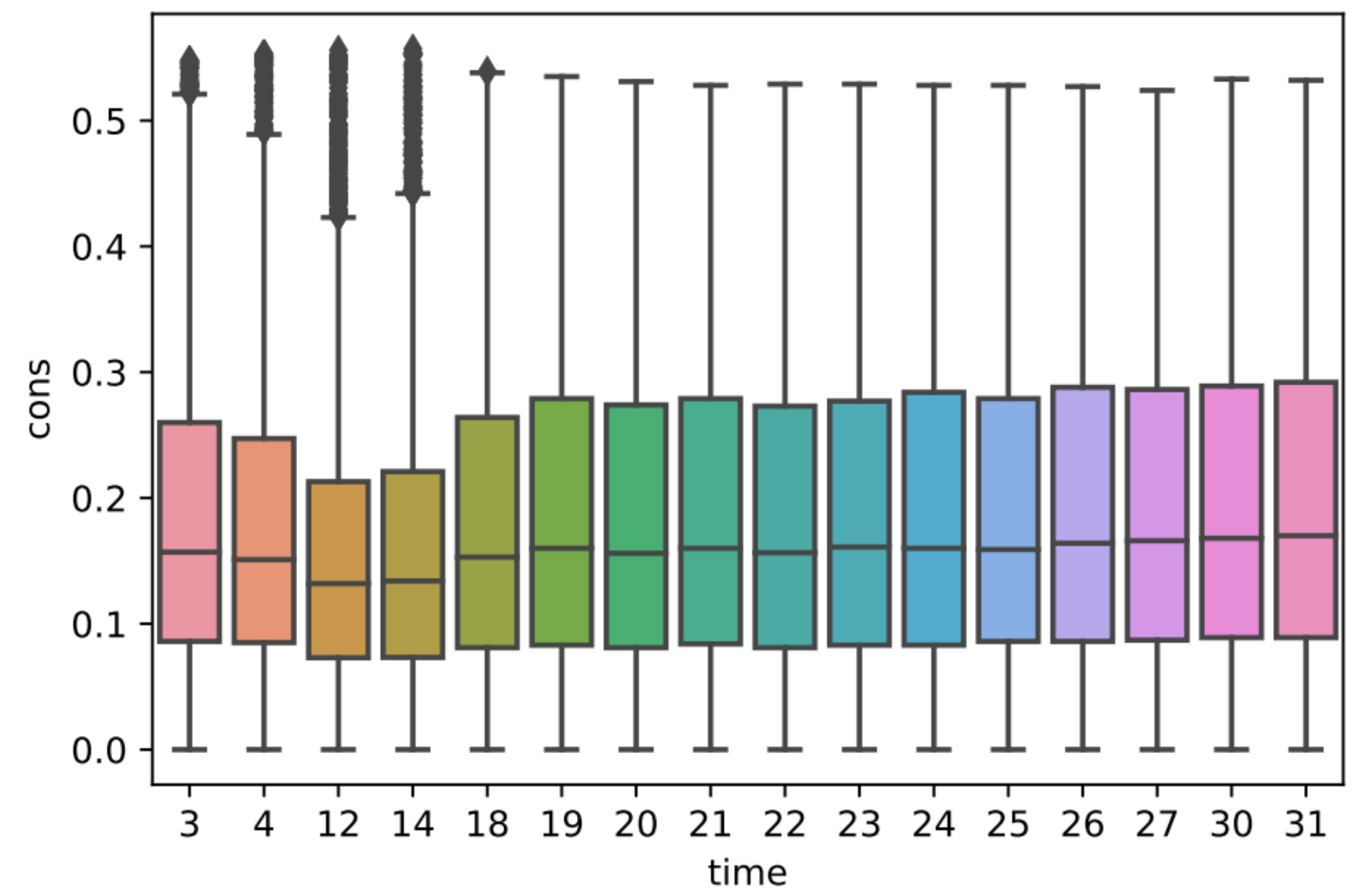
Meter ID	Day of Year	Time of Day	Energy Consumption (kW-h)
1860	54	1:30 am	0.140
1860	55	1:30 am	0.138
...	...	...	...
1860	180	3:30 pm	1.536
1860	180	4:00 pm	1.742
...	...	...	...
1610	258	1:30 am	10.536
...	...	...	...
1610	265	3:30 pm	12.647
...	...	...	...

- We collected the electricity consumption data provided by the Irish Social Science Data Archive Center.
- Each data record has three main attributes: meter ID, date/time of collection, and the energy consumption data in kW-h.

Dataset: Data Preprocessing

- The date and time are not a continuous valued attributes, rather they are categorical values, which required one-hot encoding.
- We found any missing records corresponding to any particular time, and used the average of the preceding and succeeding record to fill in the missing value.
- We consider the z−score of the consumption value according to the formula x ← (x−μ)/σ so that the variables possess approximately zero mean, which in practice, reduces computational cost while training the models.

Results



Boxplot showing the anomalous data in one of the clusters, which were labeled as fraudulent.

	ANN		SVM		K-NN		Adaboost	
Cluster	C1	C2	C1	C2	C1	C2	C1	C2
TPR	0.98	0.94	0.67	0.93	0.98	0.88	0.98	0.91
FPR	0.00007	0.001	0	0.00007	0.00006	0.0009	0.00006	0.002
Precision	0.99	0.95	1	0.99	0.99	0.96	0.99	0.93
Recall	0.98	0.95	0.67	0.93	0.98	0.88	0.98	0.91
F1	0.99	0.94	0.8	0.96	0.99	0.92	0.99	0.92
Accuracy	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99

A comparison of performance between ANN and other supervised techniques. The results are shown for two of the five clusters found from the unsupervised technique.

References

[1] "Smart meters pose security risks to consumers, utilities: Researcher," https://www.securityweek.com/smart-meters-pose-security-risks-consumers-utilities-researcher..

[2]F. Skopik, Z. Ma, T. Bleier, and H. Gruneis, "A survey on threats and vulnerabilities in smart metering infrastructures," International Journal of Smart Grid and Clean Energy, vol. 1, no. 1, pp. 22–28,2012.

[3] D. Grochocki, J. H. Huh, R. Berthier, R. Bobba, W. H. Sanders, A. A. Cardenas, and J. G. Jetcheva, "Ami threats, intrusion detection requirements and deployment recommendations," in Smart Grid Communications (SmartGridComm), 2012 IEEE Third International Conference on. IEEE, 2012, pp. 395–400.