

Name: Gaurav Kumar Singh

Registration Number: 19BCE2119

Course: CSE3501-ELA-L21+L22 (ISAA)

Lab FAT

Question set: 2a

Develop a Machine Learning based Malware detection system using any of the following Logistic Regression [30]

b. Using Cisco Packet Tracer configure **LAN** that acts as a connector between two computers to enable message transfer. [20]

1)

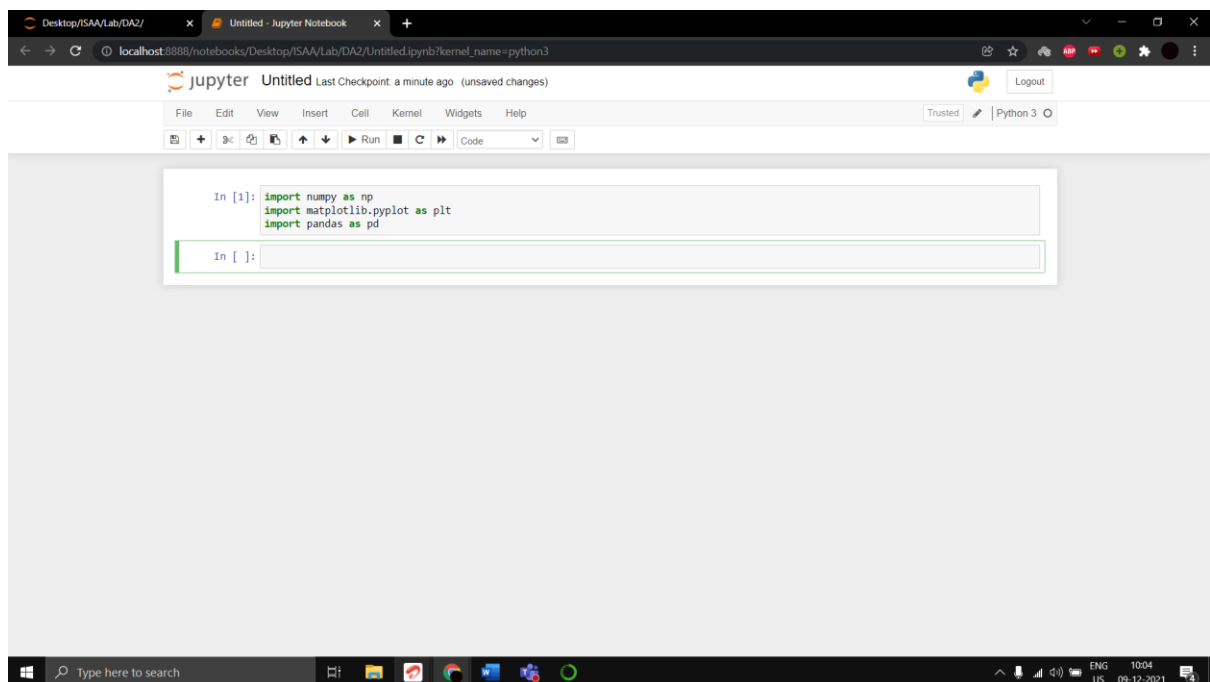
Logistic Regression (Full Code)

Importing python libraries such as NumPy, matplotlib and pandas for data reading and manipulation.

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

```
import pandas as pd
```



Reading dataset and truncating irrelevant columns

```
dataset = pd.read_csv('Malware_dataset.csv')

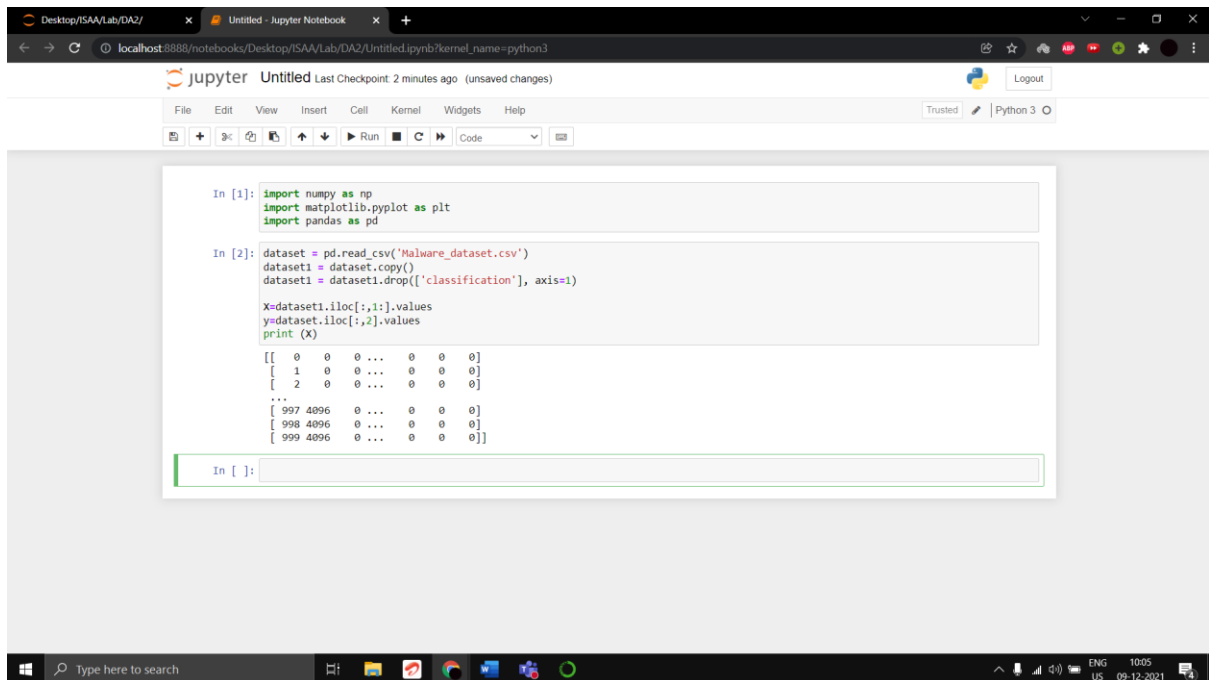
dataset1 = dataset.copy()

dataset1 = dataset1.drop(['classification'], axis=1)

X=dataset1.iloc[:,1:].values

y=dataset1.iloc[:,2].values

print (X)
```



The screenshot shows a Jupyter Notebook running in a web browser. The notebook has a single cell with the following code:

```
In [1]: import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

In [2]: dataset = pd.read_csv('Malware_dataset.csv')
dataset1 = dataset.copy()
dataset1 = dataset1.drop(['classification'], axis=1)

X=dataset1.iloc[:,1:].values
y=dataset1.iloc[:,2].values
print (X)
```

The output of the code is displayed below the cell:

```
[[ 0  0  0  0 ...  0  0  0]
 [ 1  0  0  0 ...  0  0  0]
 [ 2  0  0  0 ...  0  0  0]
 ...
 [997 4896 0 ... 0  0  0]
 [998 4896 0 ... 0  0  0]
 [999 4896 0 ... 0  0  0]]
```

The Jupyter Notebook interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help), a toolbar with icons for file operations and execution, and a status bar at the bottom showing the system clock and language settings.

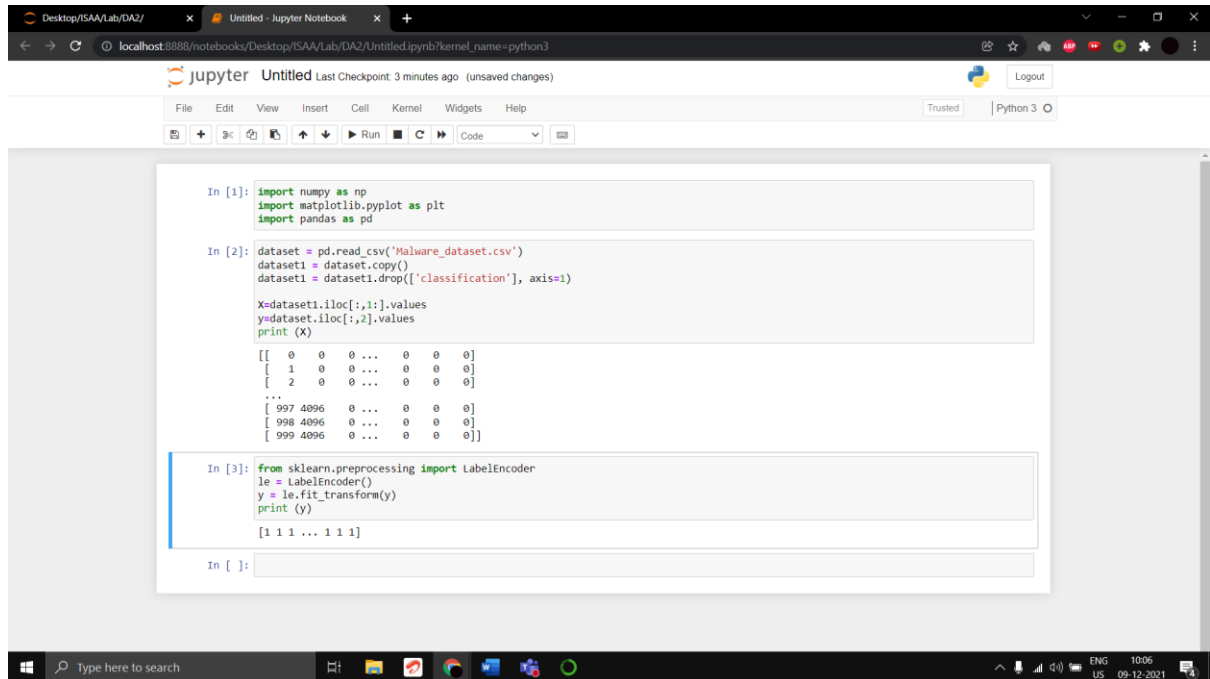
Importing sklearn.preprocessing module to transform classification dataset using label encoder for classification.

```
from sklearn.preprocessing import LabelEncoder
```

```
le = LabelEncoder()
```

```
y = le.fit_transform(y)
```

```
print (y)
```



The screenshot shows a Jupyter Notebook running in a web browser. The notebook has three code cells. The first cell imports numpy, matplotlib, and pandas. The second cell reads a CSV file, drops the 'classification' column, and prints the first few rows of the resulting dataset. The third cell imports LabelEncoder from sklearn.preprocessing, creates a LabelEncoder object, fits it to the 'classification' column, and prints the transformed labels. The output of the third cell shows the transformed labels as a list of integers.

```
In [1]: import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

In [2]: dataset = pd.read_csv('Malware_dataset.csv')
dataset1 = dataset.copy()
dataset1 = dataset1.drop(['classification'], axis=1)

X=dataset1.iloc[:,1:].values
y=dataset1.iloc[:,2].values
print (X)

[[ 0  0  0  0 ...  0  0  0]
 [ 1  0  0  0 ...  0  0  0]
 [ 2  0  0  0 ...  0  0  0]
 ...
 [997 4096  0 ...  0  0  0]
 [998 4096  0 ...  0  0  0]
 [999 4096  0 ...  0  0  0]]

In [3]: from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
y = le.fit_transform(y)
print (y)

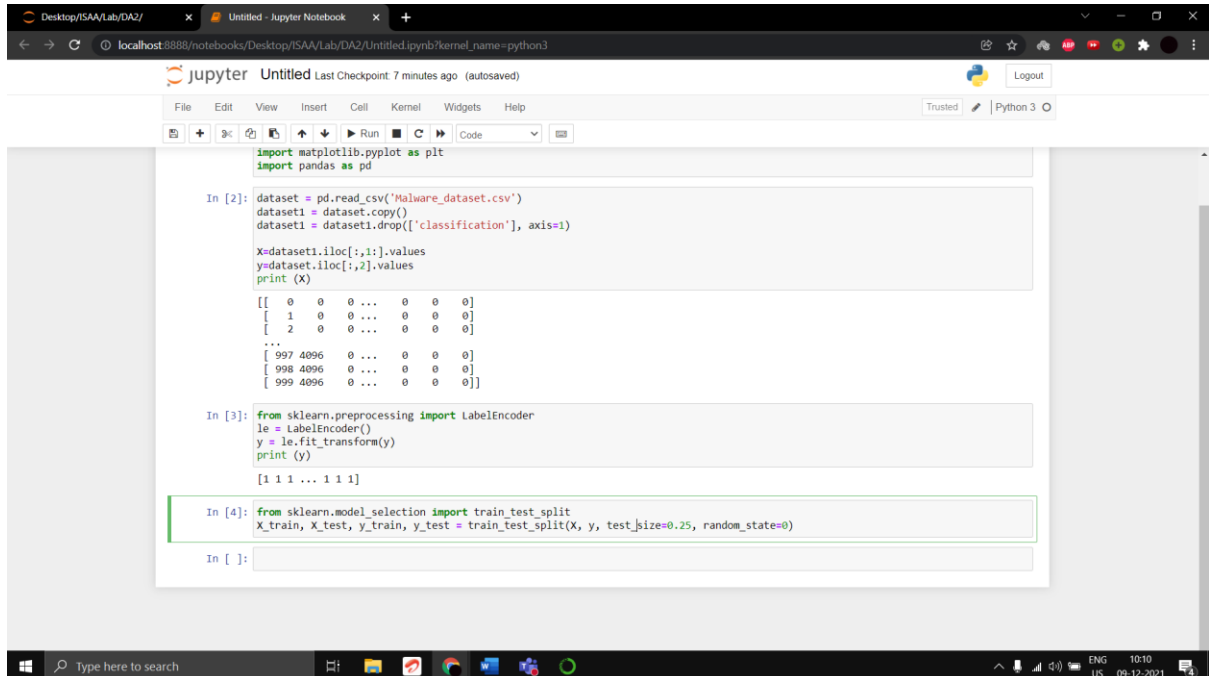
[1 1 1 ... 1 1 1]
```

Splitting dataset into training and testing dataset to calculate accuracy precision and recall on.

Test size: Train size:: 1: 3

```
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=0)
```



The screenshot shows a Jupyter Notebook window titled 'Untitled' with a Python 3 kernel. The notebook contains four code cells. The first cell imports matplotlib.pyplot as plt and pandas as pd. The second cell reads a CSV file named 'Malware_dataset.csv', copies it to dataset1, drops the 'classification' column, and prints the first few rows of the resulting X and y arrays. The third cell imports LabelEncoder from sklearn.preprocessing and fits it to the y array. The fourth cell imports train_test_split from sklearn.model_selection and splits the data into training and testing sets with a test size of 0.25 and random state 0.

```
import matplotlib.pyplot as plt
import pandas as pd

In [2]: dataset = pd.read_csv('Malware_dataset.csv')
dataset1 = dataset.copy()
dataset1 = dataset1.drop(['classification'], axis=1)

X=dataset1.iloc[:,1:].values
y=dataset1.iloc[:,2].values
print(X)

[[ 0  0  0  0 ...  0  0  0]
 [ 1  0  0  0 ...  0  0  0]
 [ 2  0  0  0 ...  0  0  0]
 ...
 [997 4096 0 ...  0  0  0]
 [998 4096 0 ...  0  0  0]
 [999 4096 0 ...  0  0  0]]

In [3]: from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
y = le.fit_transform(y)
print(y)

[1 1 1 ... 1 1 1]

In [4]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=0)

In [ ]:
```

Printing the training and testing datasets

```
print("Xtrain")
```

```
print(X_train)
```

```
print("Xtest")
```

```
print(X_test)
```

```
print("ytrain")
```

```
print(y_train)
```

```
print("ytest")
```

```
print(y_test)
```

```
Desktop/ISAA/Lab/DA2/ x Untitled - Jupyter Notebook x +
localhost:8888/notebooks/Desktop/ISAA/Lab/DA2/Untitled.ipynb?kernel_name=python3

jupyter Untitled Last Checkpoint: 11 minutes ago (unsaved changes)
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

[1 1 1 ... 1 1 1]

In [4]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=0)

In [7]: print("Xtrain")
print(X_train)
print("Xtest")
print(X_test)
print("ytrain")
print(y_train)
print("ytest")
print(y_test)

Xtrain
[[ 606 12288  0 ... 10  0  0]
 [ 228 28672  0 ...  0  0  0]
 [ 382 4096  0 ...  1  0  0]
 ...
 [ 613 12288  0 ... 11  0  0]
 [ 567 4096  0 ...  2  0  0]
 [ 268  0  0 ...  0  0  0]]

Xtest
[[ 582  0  0 ...  8  0  0]
 [ 498  0  0 ...  0  0  0]
 [ 227 1028096  0 ...  4  0  0]
 ...
 [ 585 4096  0 ...  0  0  0]
 [ 519  0  0 ...  7  0  0]
 [ 831  0  0 ...  0  0  0]]

ytrain
[0 1 0 ... 0 0 1]
ytest
[1 0 0 ... 1 0 0]

In [ ]:
```

Scaler transform of the training dataset

```
from sklearn.preprocessing import StandardScaler
```

```
sc= StandardScaler()
```

```
X_train= sc.fit_transform(X_train)
```

```
X_test= sc.transform(X_test)
```

```
print(X_train)
```

```
Desktop/ISAA/Lab/DA2/ x Untitled - Jupyter Notebook x +
localhost:8888/notebooks/Desktop/ISAA/Lab/DA2/Untitled.ipynb?kernel_name=python3

jupyter Untitled Last Checkpoint: 12 minutes ago (autosaved)
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

...
[ 585 4096  0 ...  0  0  0]
[ 519  0  0 ...  7  0  0]
[ 831  0  0 ...  0  0  0]]

ytrain
[0 1 0 ... 0 0 1]
ytest
[1 0 0 ... 1 0 0]

In [8]: from sklearn.preprocessing import StandardScaler
sc= StandardScaler()
X_train= sc.fit_transform(X_train)
X_test= sc.transform(X_test)
print(X_train)

[[ 0.36731703 -0.15969785  0. ... 2.5559669  0.
  0.
 [-0.941068 -0.14179605  0. ... -0.50896701  0.
  0.
 [-0.40802225 -0.16864875  0. ... -0.20247362  0.
  0.
 ...
 [ 0.39154638 -0.15969785  0. ... 2.8624603  0.
  0.
 [ 0.23232492 -0.16864875  0. ... 0.10401977  0.
  0.
 [-0.80261455 -0.1731242  0. ... -0.50896701  0.
  0.
]]

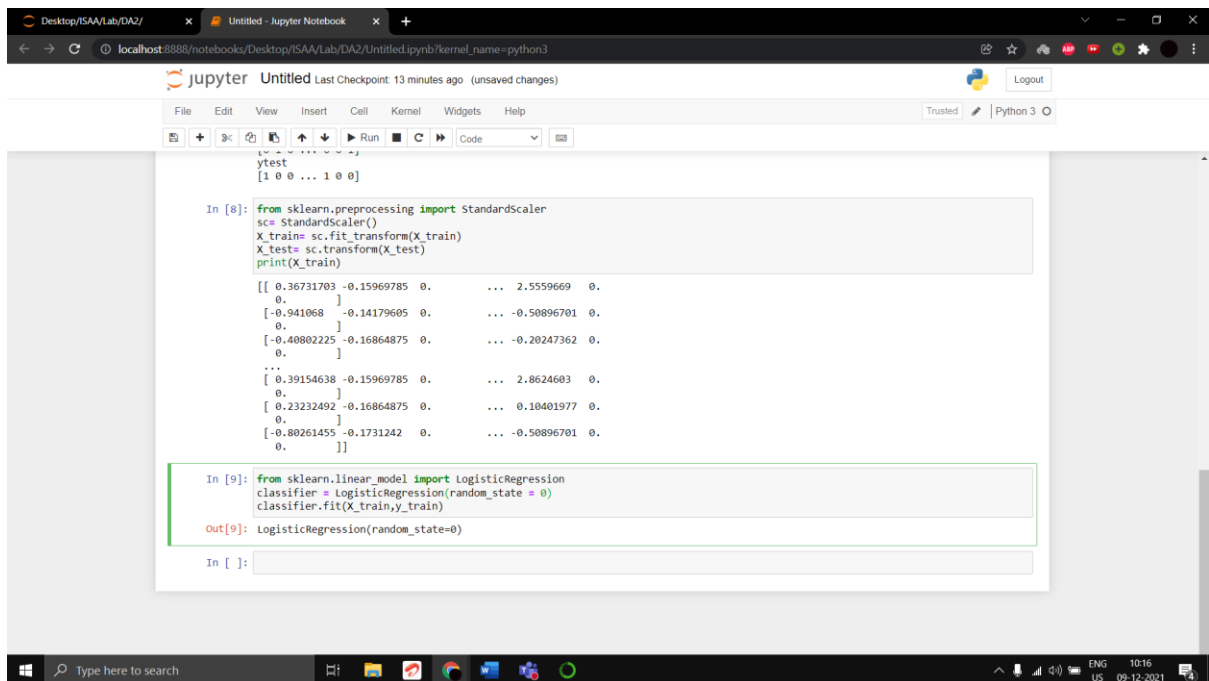
In [ ]:
```

Training the algorithm using Logistic regression

```
from sklearn.linear_model import LogisticRegression
```

```
classifier = LogisticRegression(random_state = 0)
```

```
classifier.fit(X_train,y_train)
```



```
Desktop/ISAA/Lab/DA2/ x Untitled - Jupyter Notebook x +
localhost:8888/notebooks/Desktop/ISAA/Lab/DA2/Untitled.ipynb?kernel_name=python3
jupyter Untitled Last Checkpoint: 13 minutes ago (unsaved changes) Logout
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3
ytest
[1 0 0 ... 1 0 0]

In [8]: from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)
print(X_train)

[[ 0.36731703 -0.15969785  0.         ... 2.5559669  0.
  0.         ]
 [-0.941068   -0.14179605  0.         ... -0.50896701  0.
  0.         ]
 [-0.40802225 -0.16864875  0.         ... -0.20247362  0.
  0.         ]
 ...
 [ 0.39154638 -0.15969785  0.         ... 2.8624603  0.
  0.         ]
 [ 0.23232492 -0.16864875  0.         ... 0.10401977  0.
  0.         ]
 [-0.80261455 -0.1731242  0.         ... -0.50896701  0.
  0.         ]]

In [9]: from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression(random_state = 0)
classifier.fit(X_train,y_train)

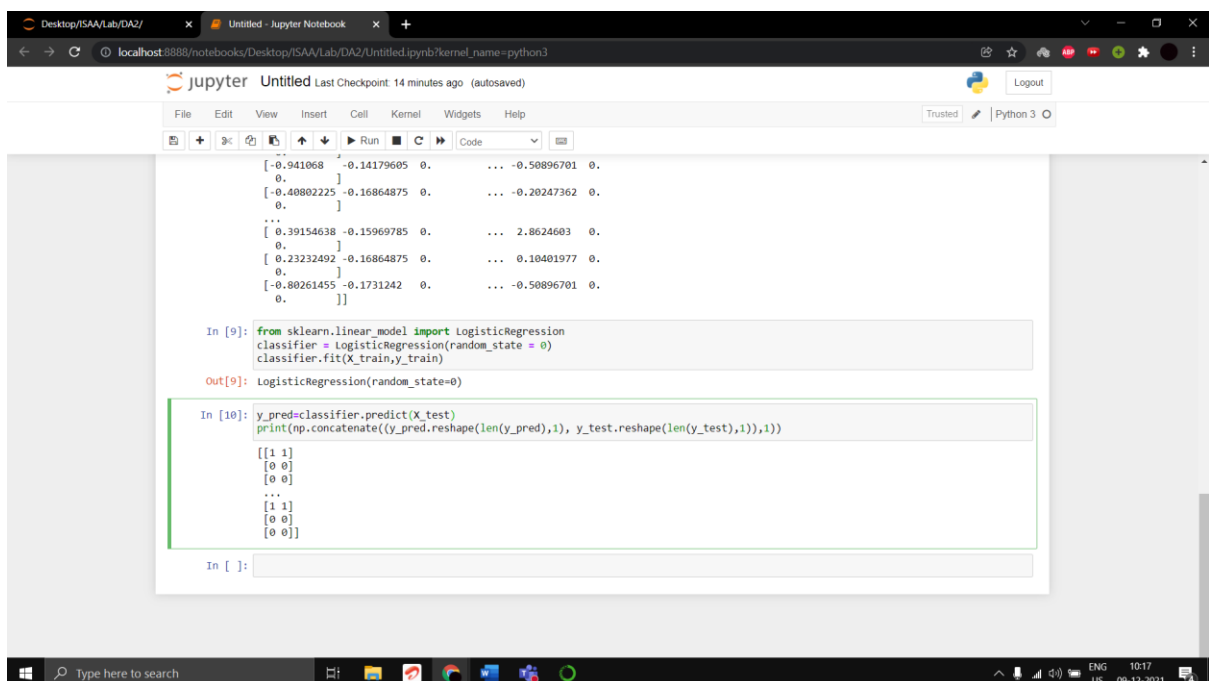
Out[9]: LogisticRegression(random_state=0)

In [ ]:
```

Testing the algorithm on the test split dataset

```
y_pred=classifier.predict(X_test)
```

```
print(np.concatenate((y_pred.reshape(len(y_pred),1), y_test.reshape(len(y_test),1)),1))
```



```
Desktop/ISAA/Lab/DA2/ x Untitled - Jupyter Notebook x +
localhost:8888/notebooks/Desktop/ISAA/Lab/DA2/Untitled.ipynb?kernel_name=python3
jupyter Untitled Last Checkpoint: 14 minutes ago (autosaved) Logout
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

[-0.941068   -0.14179605  0.         ... -0.50896701  0.
  0.         ]
[-0.40802225 -0.16864875  0.         ... -0.20247362  0.
  0.         ]
...
[ 0.39154638 -0.15969785  0.         ... 2.8624603  0.
  0.         ]
[ 0.23232492 -0.16864875  0.         ... 0.10401977  0.
  0.         ]
[-0.80261455 -0.1731242  0.         ... -0.50896701  0.
  0.         ]]

In [9]: from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression(random_state = 0)
classifier.fit(X_train,y_train)

Out[9]: LogisticRegression(random_state=0)

In [10]: y_pred=classifier.predict(X_test)
print(np.concatenate((y_pred.reshape(len(y_pred),1), y_test.reshape(len(y_test),1)),1))

[[1 1]
 [0 0]
 [0 0]
 ...
 [1 1]
 [0 0]
 [0 0]]

In [ ]:
```

Accuracy Score

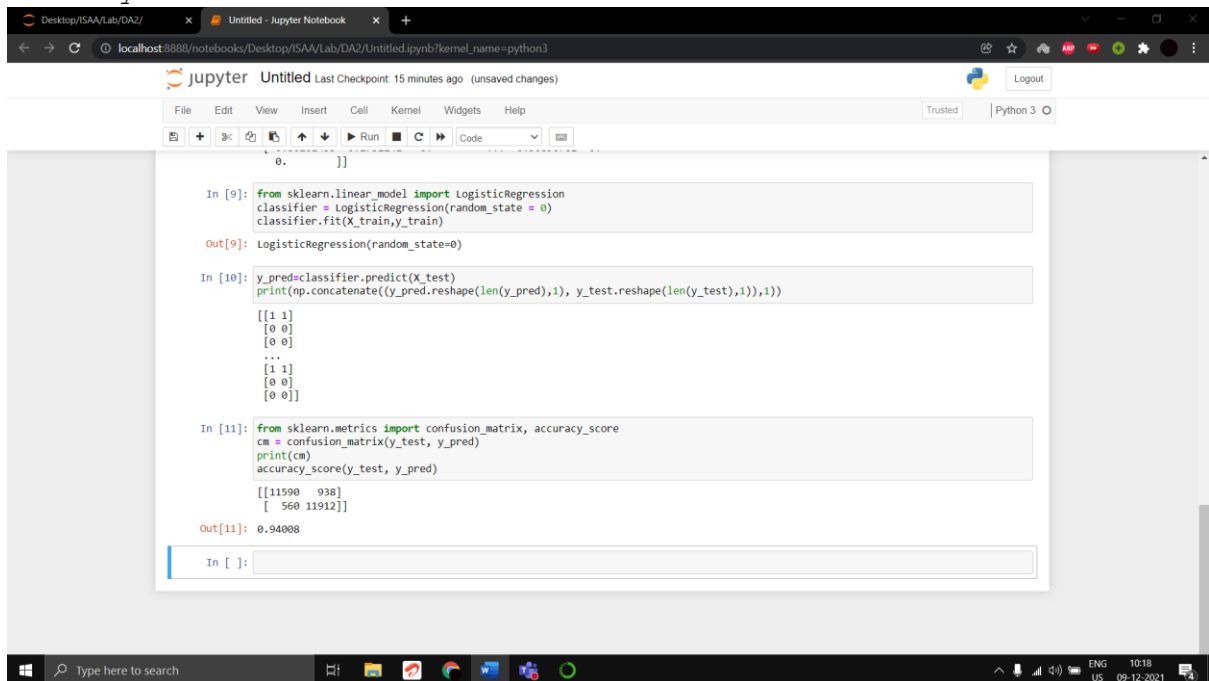
```
from sklearn.metrics import confusion_matrix, accuracy_score

cm = confusion_matrix(y_test, y_pred)

print(cm)

accuracy_score(y_test, y_pred)
```

Accuracy Score: 0.94008



Precision Score

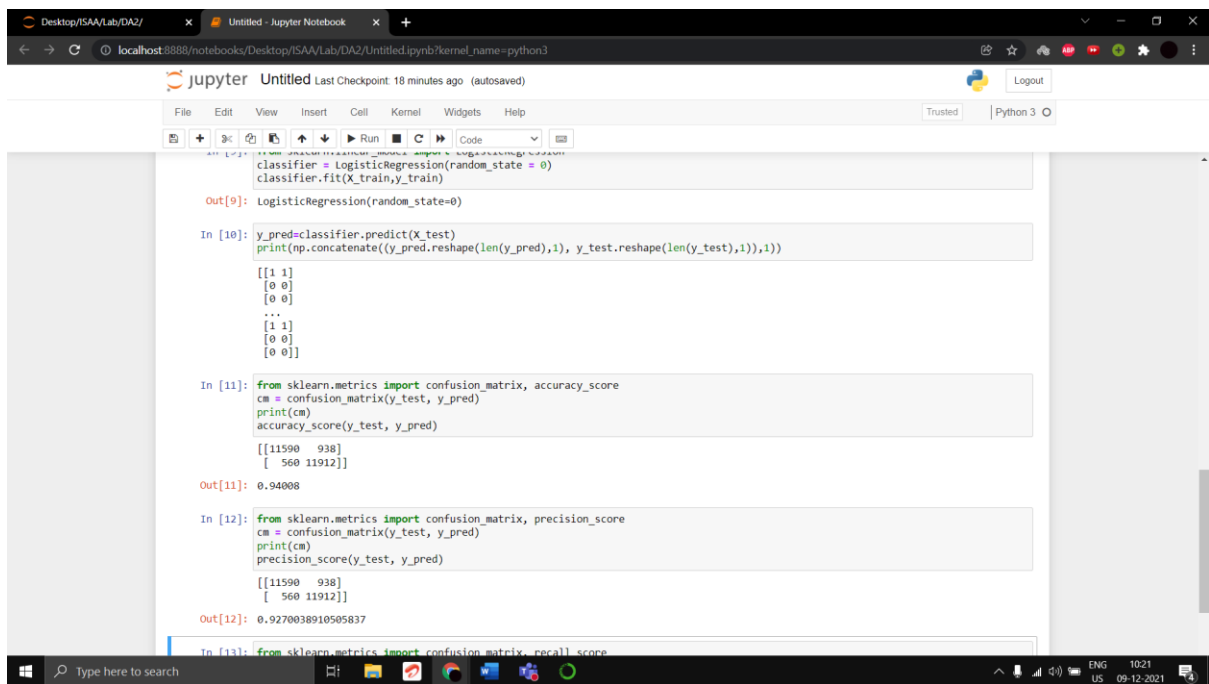
```
from sklearn.metrics import confusion_matrix, precision_score

cm = confusion_matrix(y_test, y_pred)

print(cm)

precision_score(y_test, y_pred)
```

Precision Score: 0.9270038910505837



```
from sklearn.metrics import confusion_matrix, accuracy_score
classifier = LogisticRegression(random_state = 0)
classifier.fit(X_train,y_train)

Out[9]: LogisticRegression(random_state=0)

In [10]: y_pred=classifier.predict(X_test)
print(np.concatenate((y_pred.reshape(len(y_pred),1), y_test.reshape(len(y_test),1)),1))

[[1 1]
 [0 0]
 [0 0]
 ...
 [1 1]
 [0 0]
 [0 0]]

In [11]: from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(y_test, y_pred)
print(cm)
accuracy_score(y_test, y_pred)

[[11590  938]
 [ 560 11912]]

Out[11]: 0.94008

In [12]: from sklearn.metrics import confusion_matrix, precision_score
cm = confusion_matrix(y_test, y_pred)
print(cm)
precision_score(y_test, y_pred)

[[11590  938]
 [ 560 11912]]

Out[12]: 0.9270038910505837

In [13]: from sklearn.metrics import confusion_matrix, recall_score
```

Recall Score

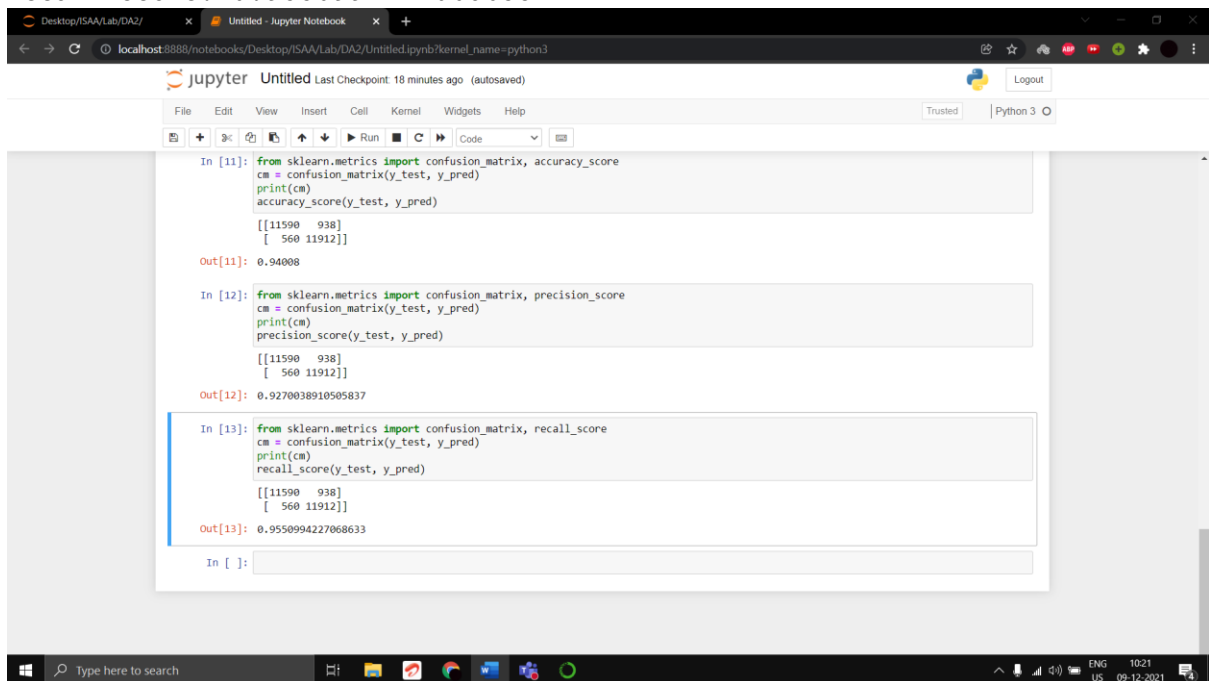
```
from sklearn.metrics import confusion_matrix, recall_score
```

```
cm = confusion_matrix(y_test, y_pred)
```

```
print(cm)
```

```
recall_score(y_test, y_pred)
```

Recall Score: 0.9550994227068633



```
from sklearn.metrics import confusion_matrix, recall_score
cm = confusion_matrix(y_test, y_pred)
print(cm)
recall_score(y_test, y_pred)

[[11590  938]
 [ 560 11912]]

Out[13]: 0.9550994227068633

In [ ]:
```

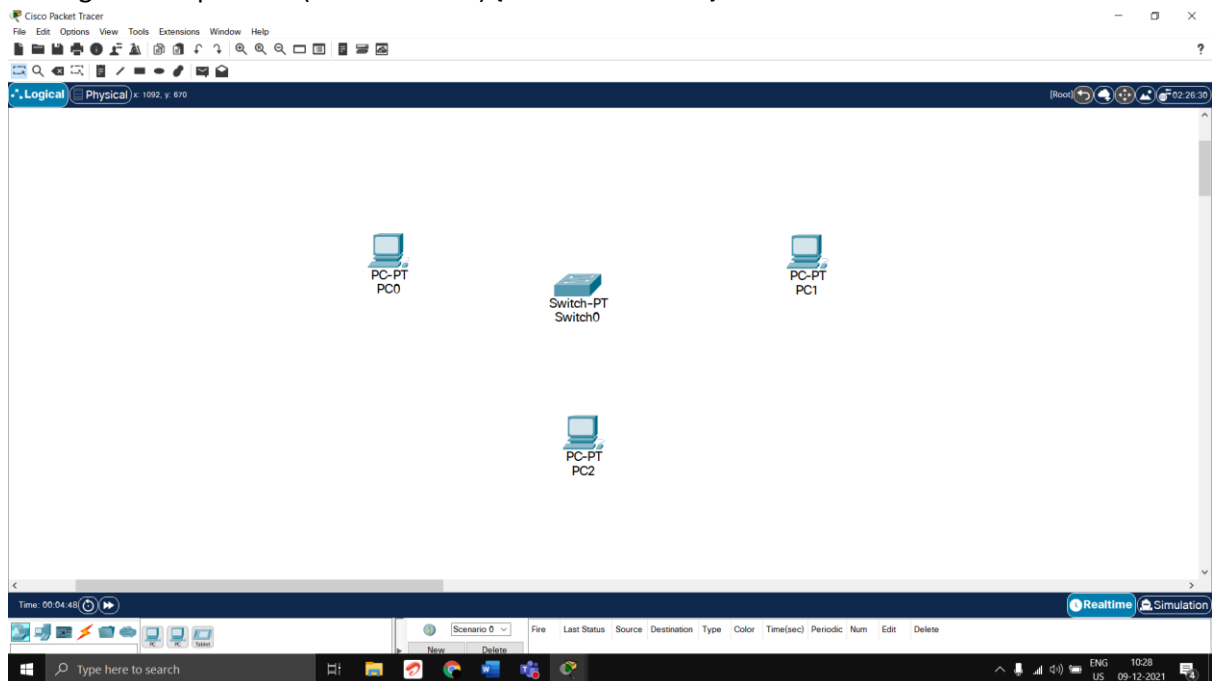

2)

Components used: PC

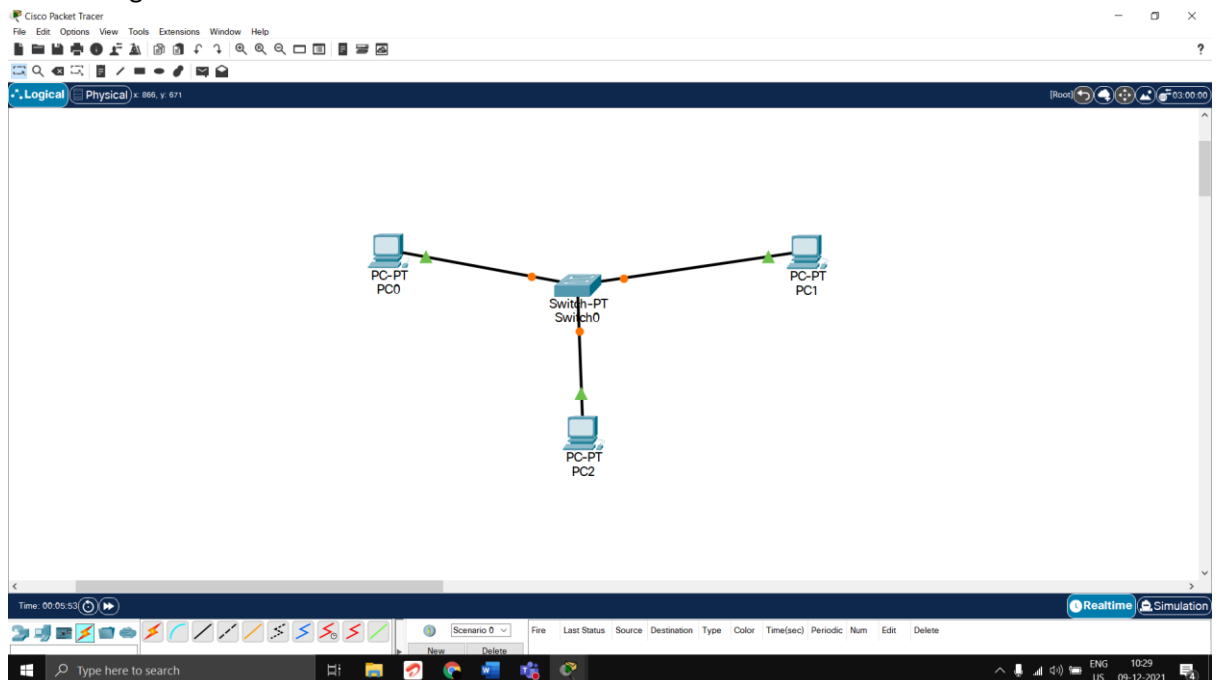
Switch

STEPS:

1) Placing the components (PC and Switch) {PC and PT-Switch}

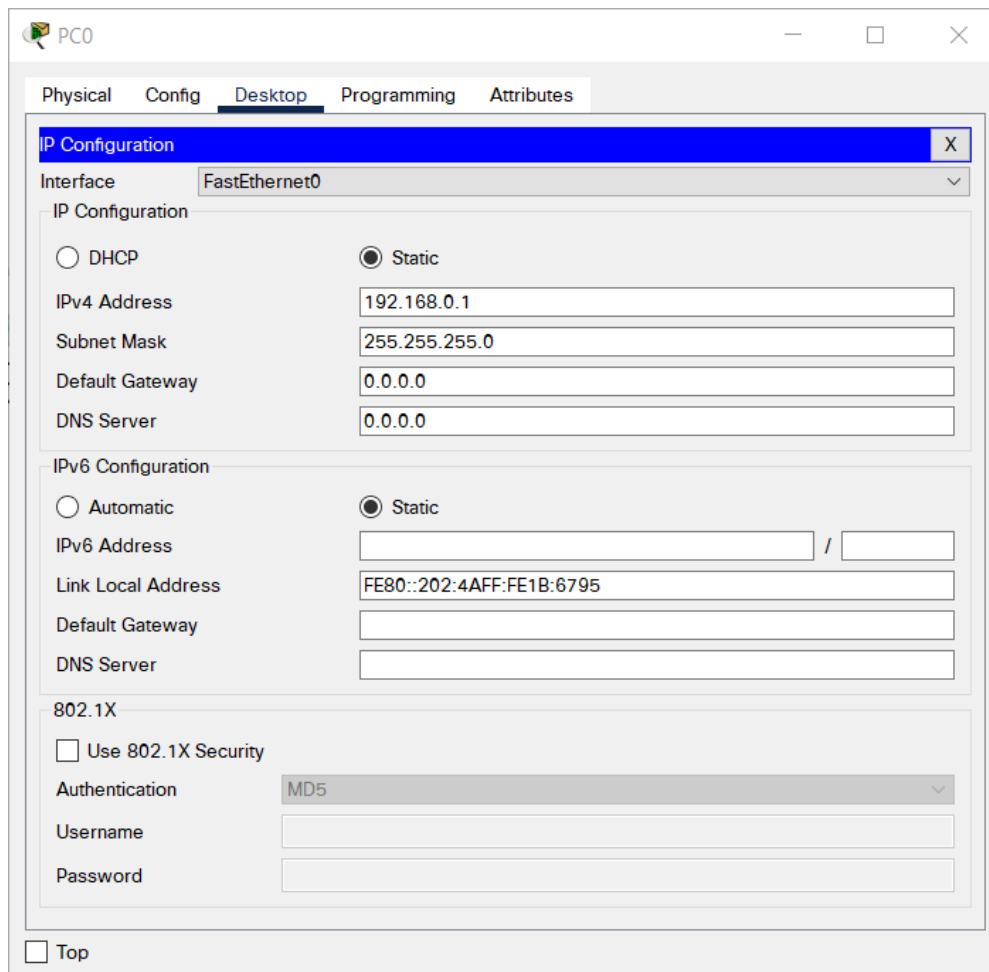
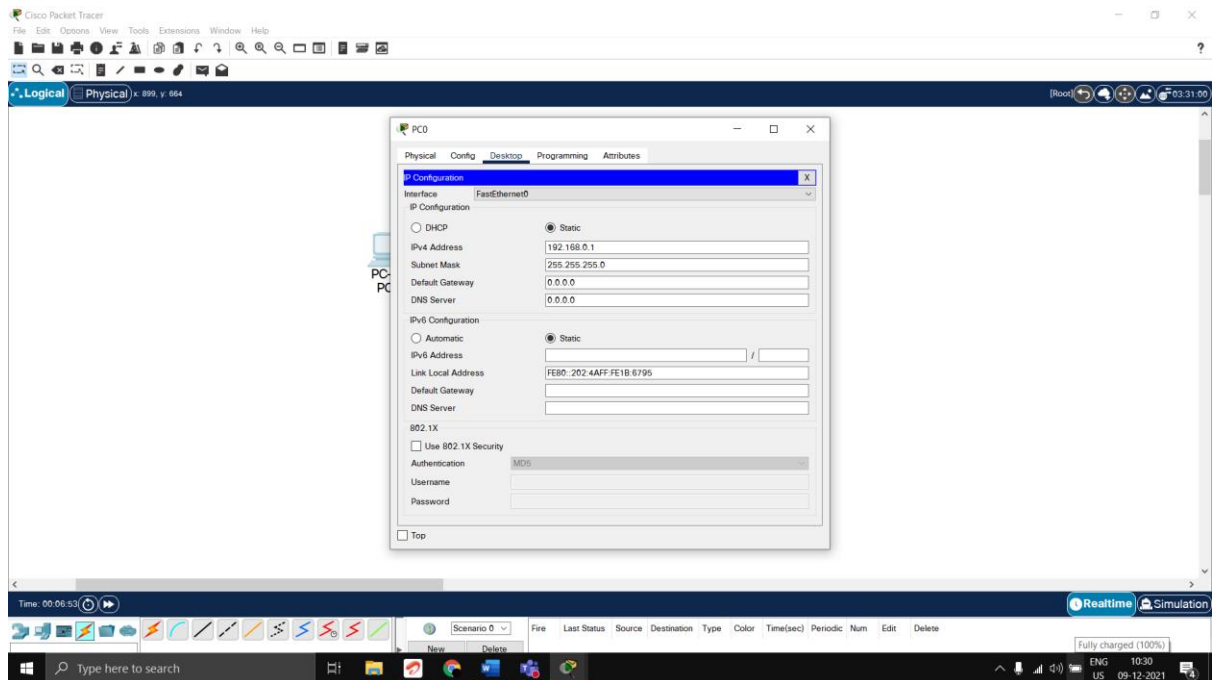


2) Connecting PCs to the Switch

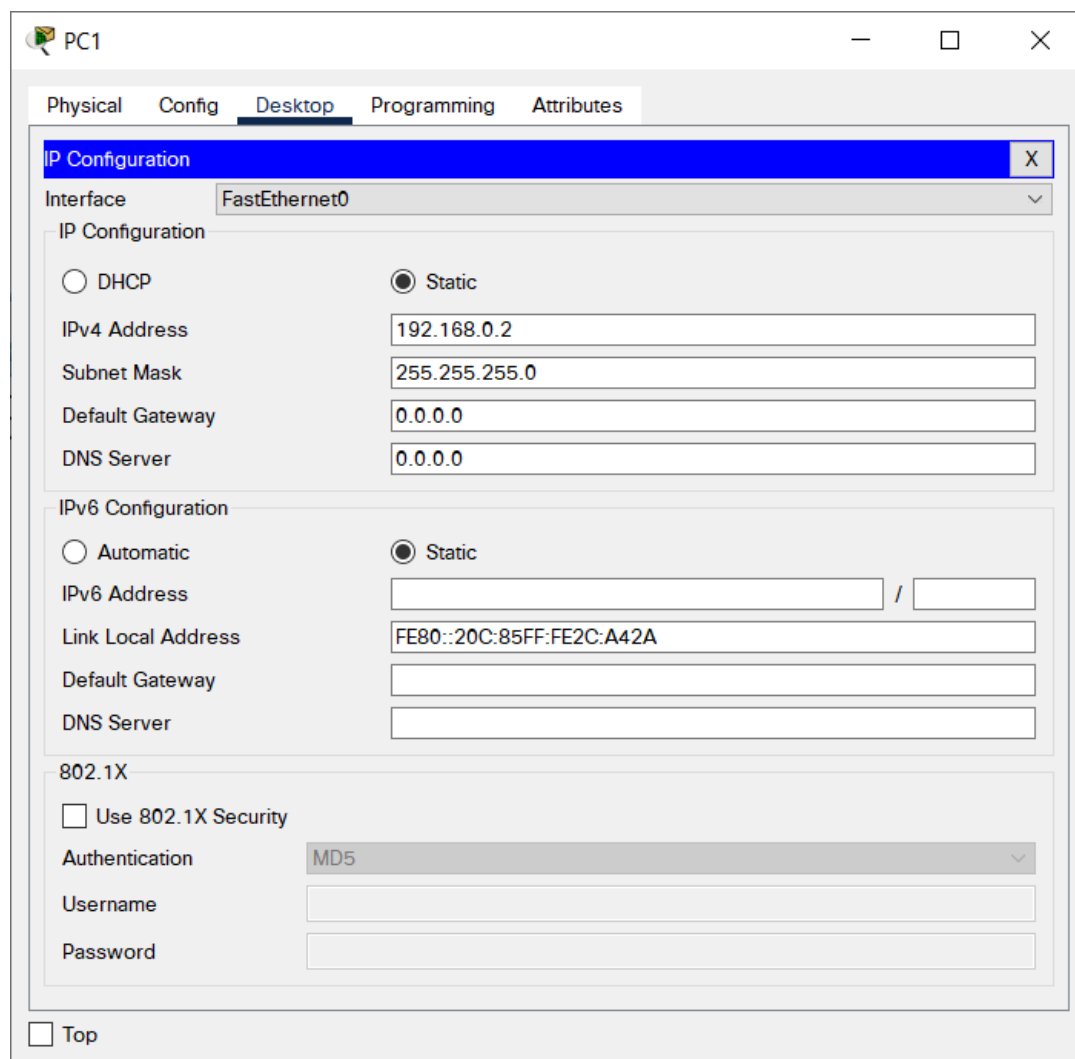
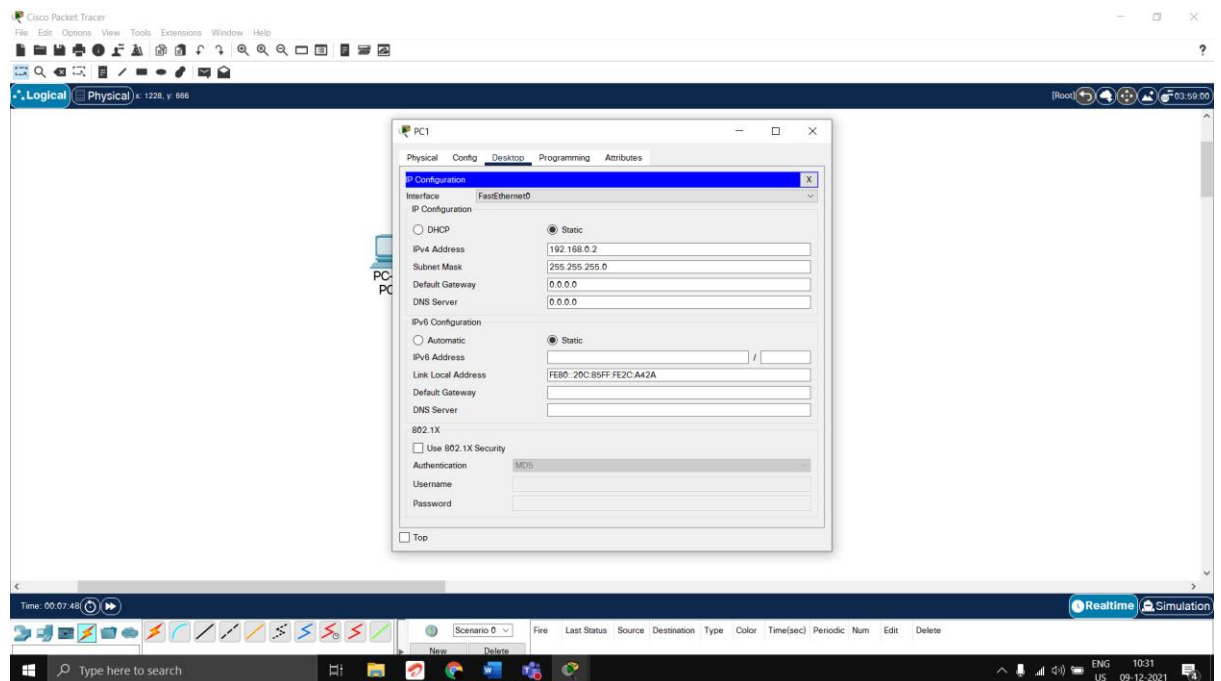


3) Configuring IP addresses and Masks of the PCs

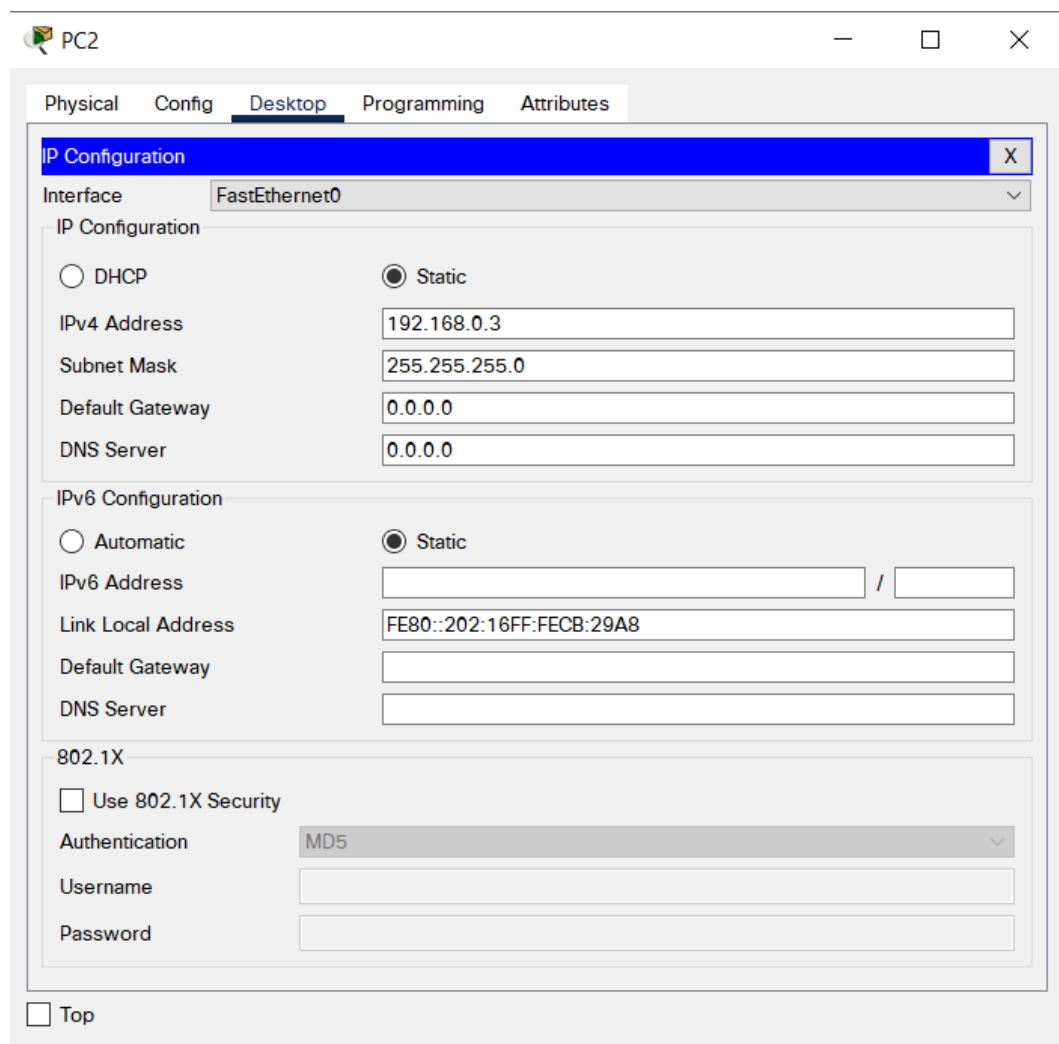
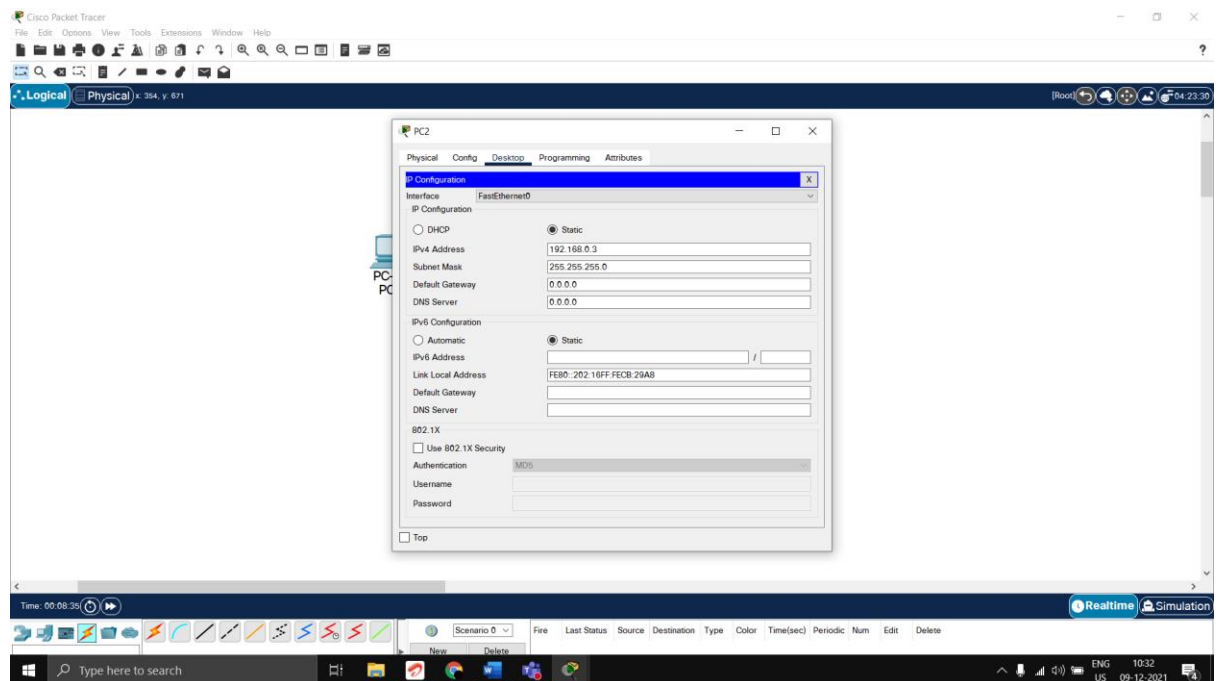
PC0



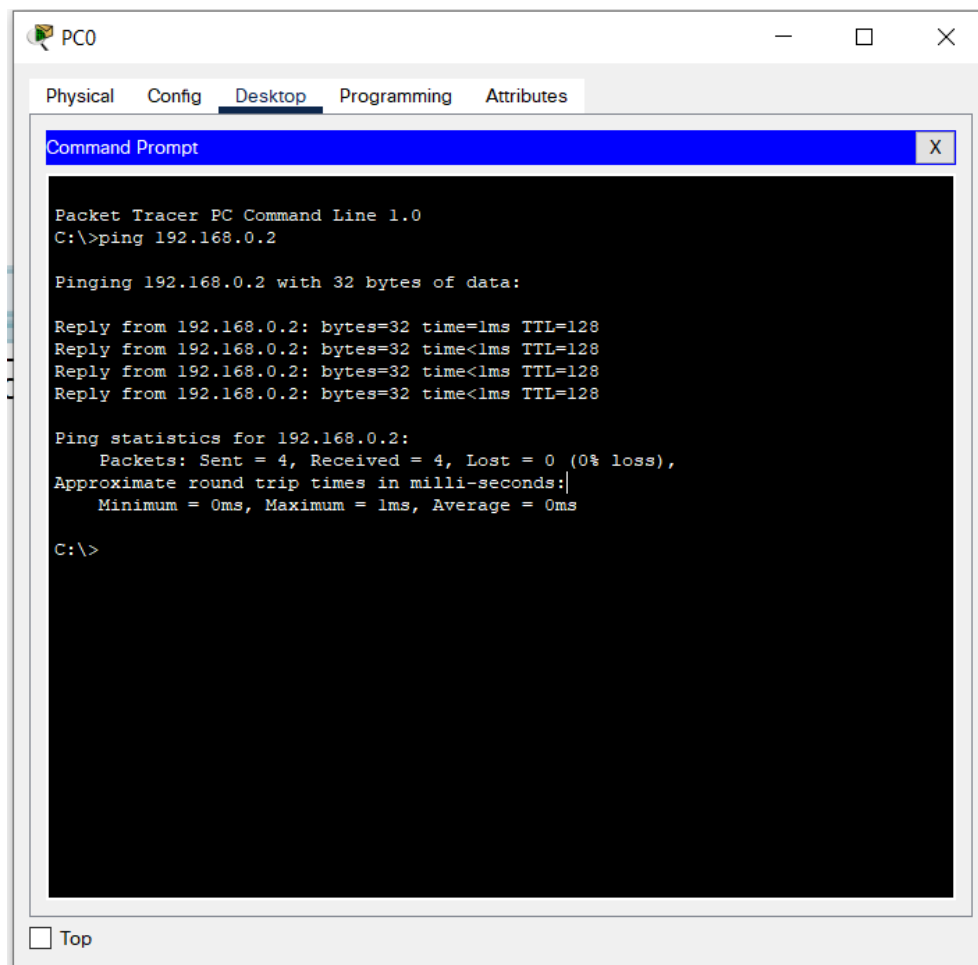
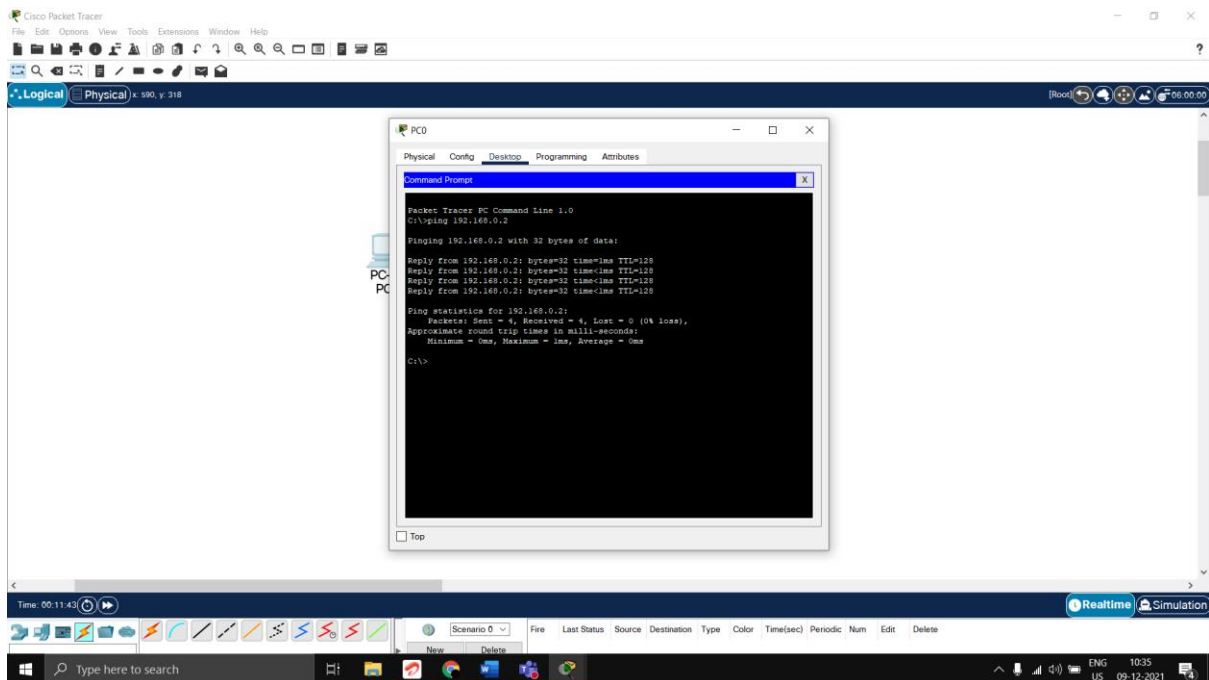
PC1



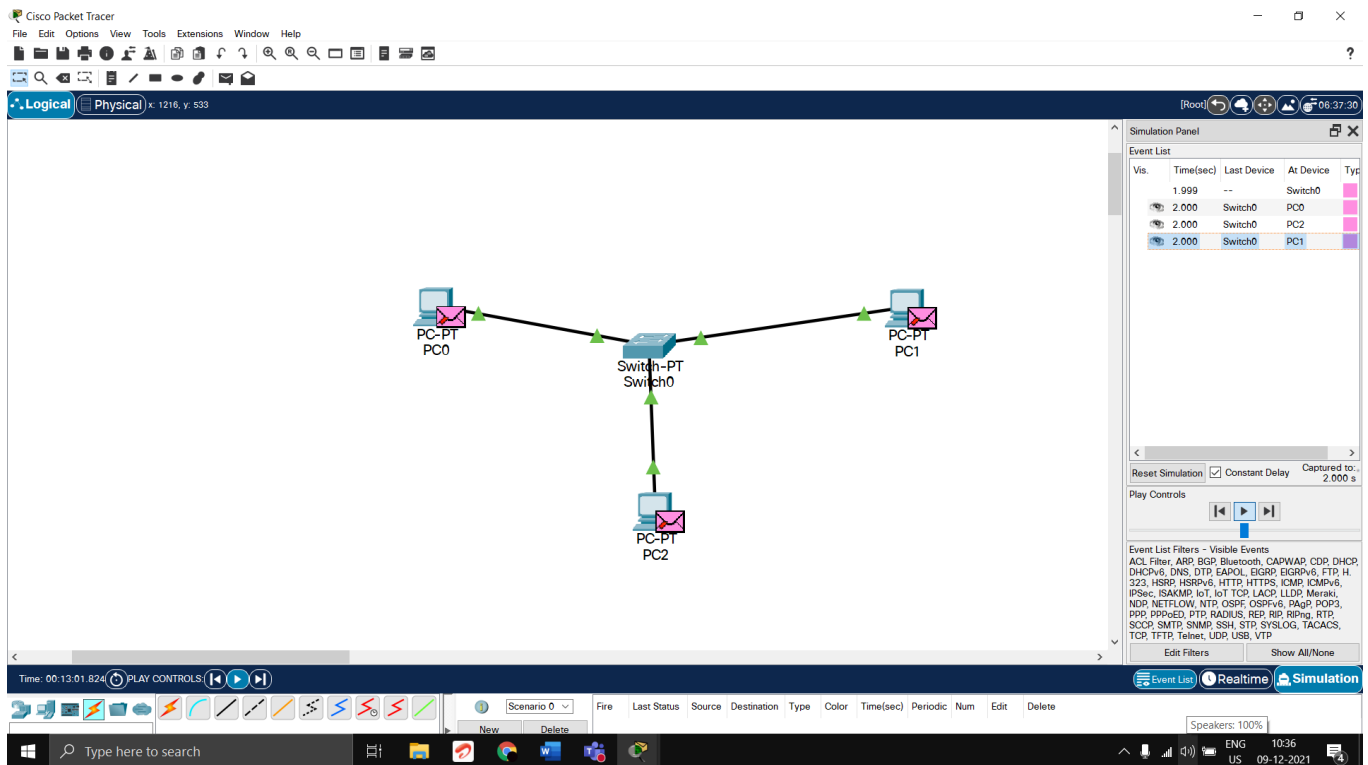
PC2



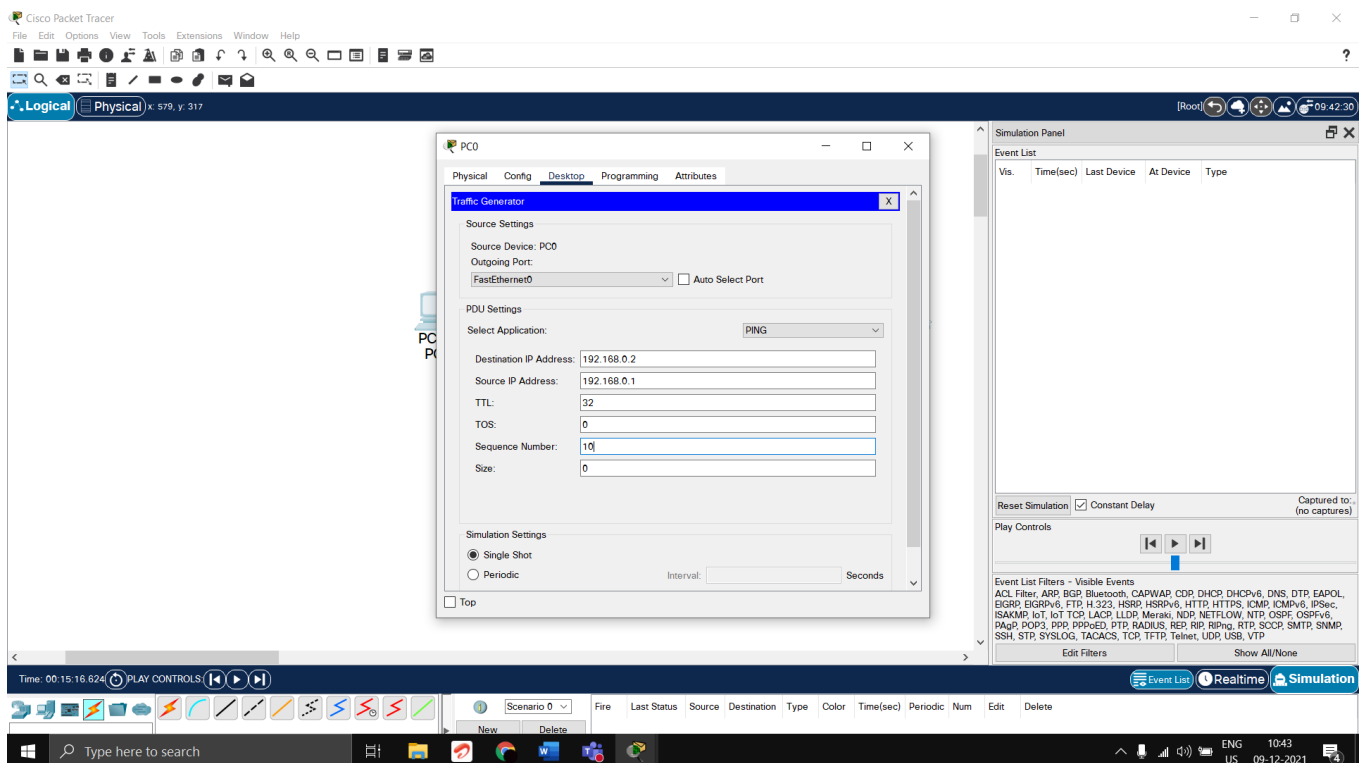
- 4) Checking the configuration using ping command and simulation
- a) Ping Command (Using PC0 to ping PC1)
- Command: ping 192.168.0.2

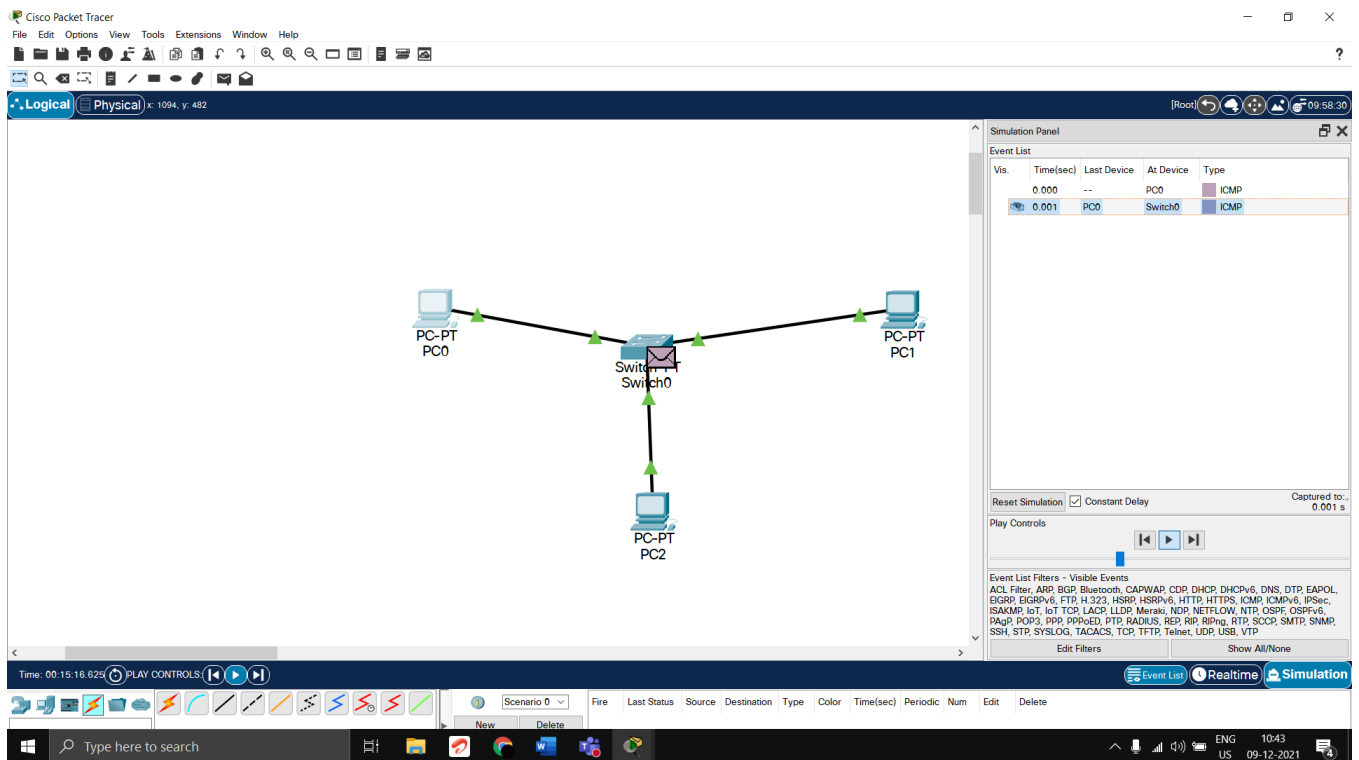
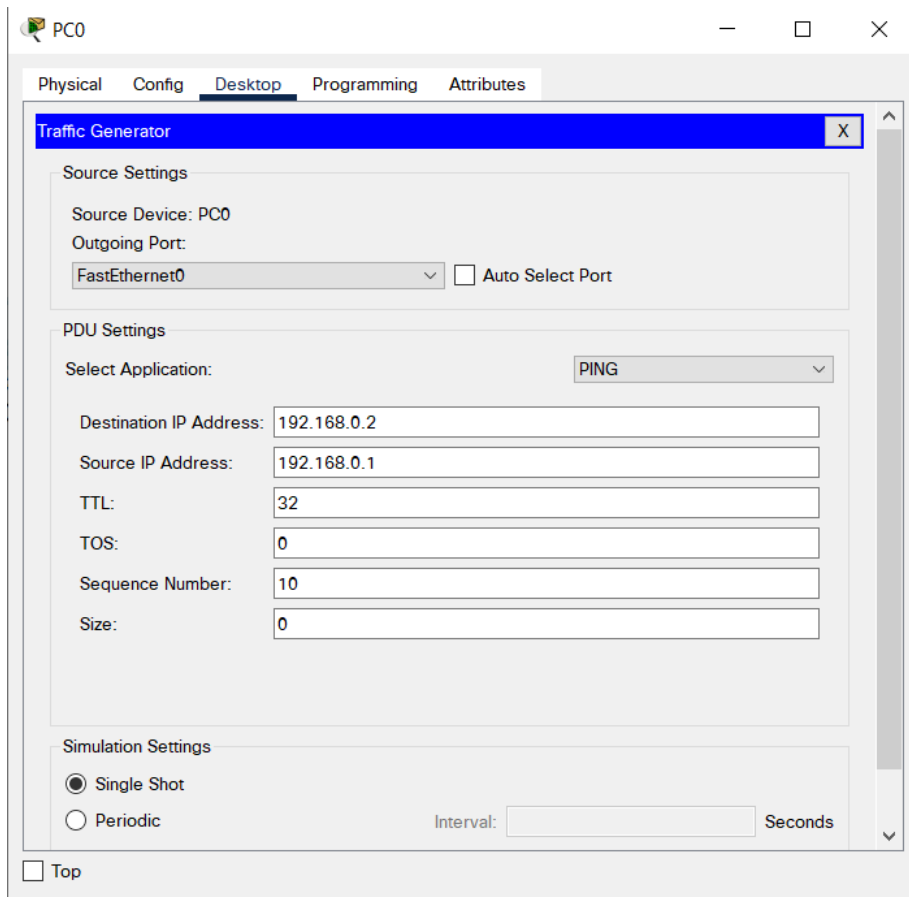


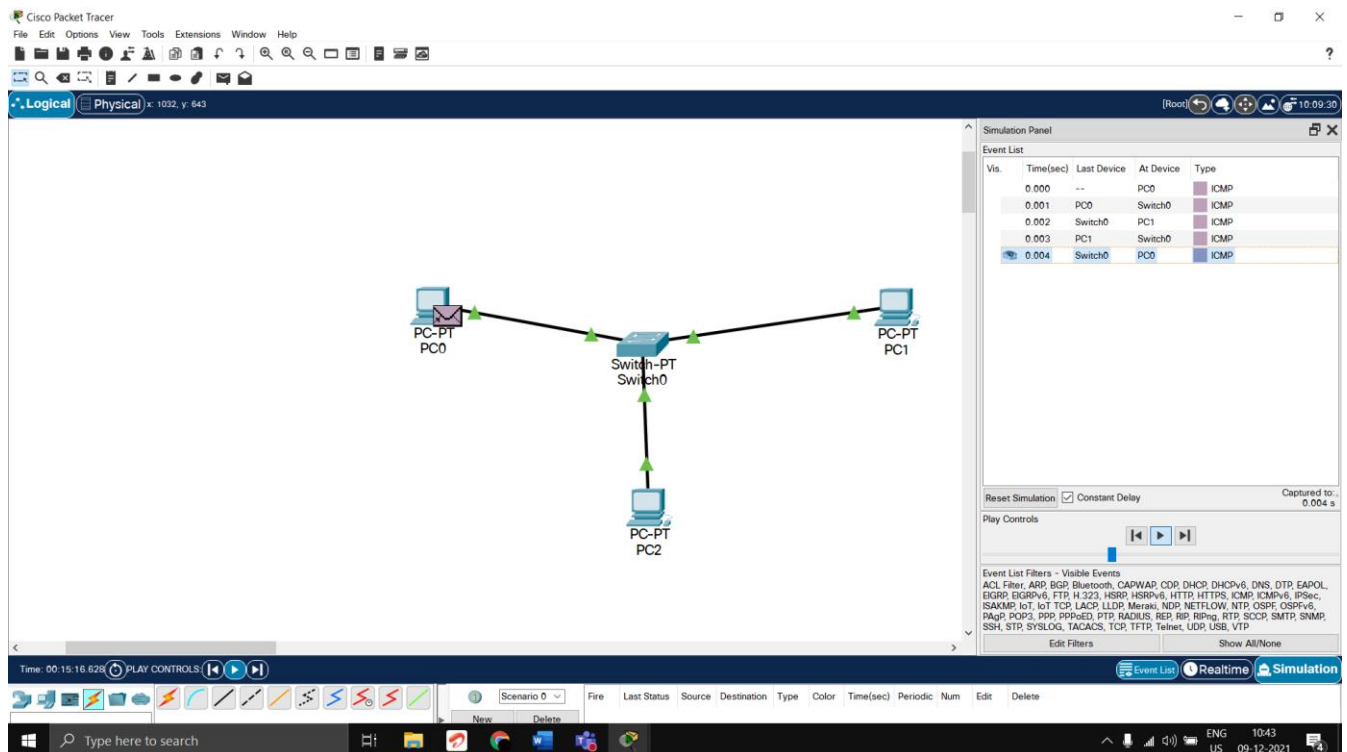
b) Using Simulation



Configuring and adding a packet to be sent during simulation







QUESTION 2a FULL CODE

```
import numpy as np

import matplotlib.pyplot as plt

import pandas as pd

dataset = pd.read_csv('Malware_dataset.csv')

dataset1 = dataset.copy()

dataset1 = dataset1.drop(['classification'], axis=1)

X=dataset1.iloc[:,1:].values

y=dataset1.iloc[:,2].values

print (X)

from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()

y = le.fit_transform(y)

print (y)

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=0)

print("Xtrain")
```



```
print(X_train)
print("Xtest")
print(X_test)
print("ytrain")
print(y_train)
print("ytest")
print(y_test)

from sklearn.preprocessing import StandardScaler
sc= StandardScaler()
X_train= sc.fit_transform(X_train)
X_test= sc.transform(X_test)
print(X_train)

from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression(random_state = 0)
classifier.fit(X_train,y_train)
y_pred=classifier.predict(X_test)
print(np.concatenate((y_pred.reshape(len(y_pred),1), y_test.reshape(len(y_test),1)),1))

from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(y_test, y_pred)
print(cm)
accuracy_score(y_test, y_pred)

from sklearn.metrics import confusion_matrix, precision_score
cm = confusion_matrix(y_test, y_pred)
print(cm)
precision_score(y_test, y_pred)

from sklearn.metrics import confusion_matrix, recall_score
cm = confusion_matrix(y_test, y_pred)
print(cm)
recall_score(y_test, y_pred)
```