Project Title

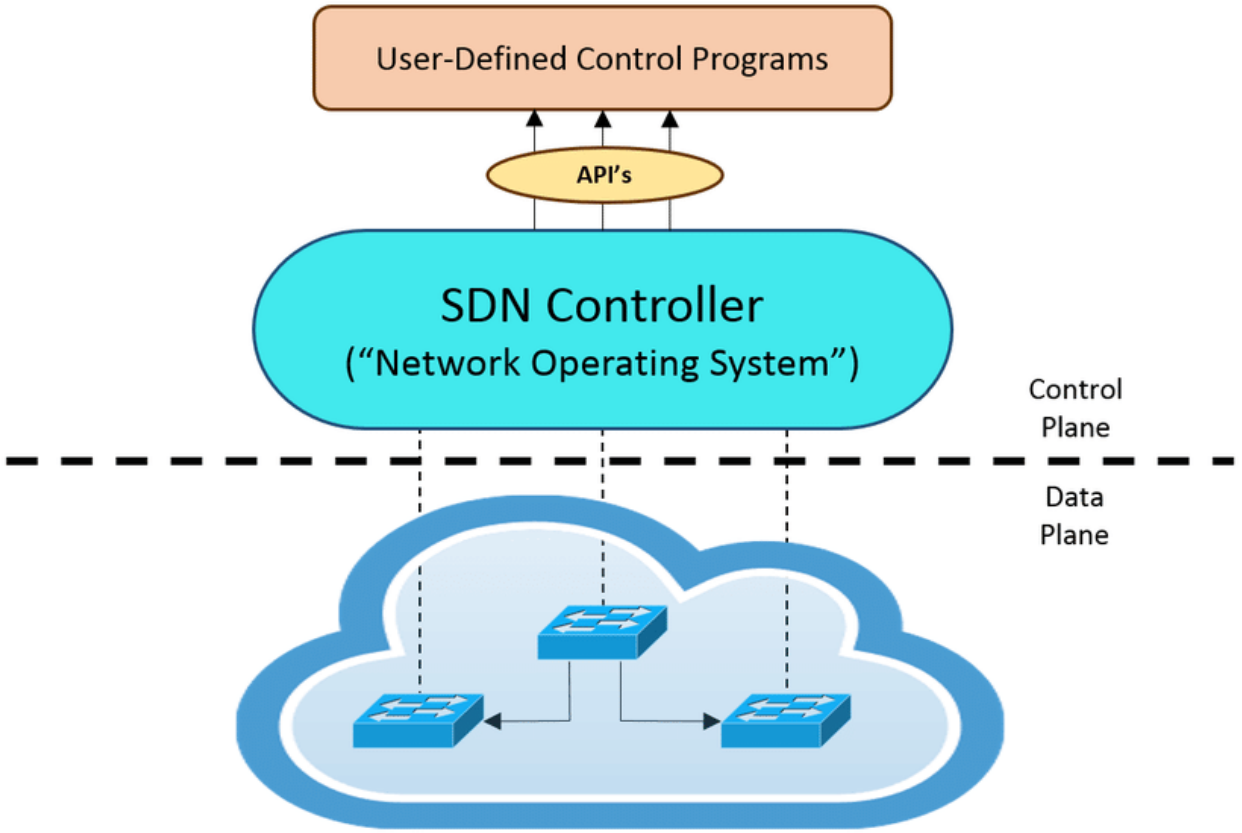
Software Defined Network (SDN) Intrusion Detection Using Machine Learning

Name	ID
Muhammad Zawad Mahmud	1931401042
Samiha Islam	1931393642
Md. Solayman Hossain	1931565042

Background

- Software-defined networking (SDN) is a networking model where software-based controllers are used to connect to network's underlying hardware architecture and control traffic.
- It is one of the most popular models currently. As a result hackers try to take control of it to gain confidential information.
- In 2000, Yahoo was the first victim of an attack and in the same date also it recorded its first ever attack publicly. At the present time, web services and social websites are target of this attackers.
- Machine learning, a form of artificial intelligence (AI) can be used to predict outcomes very accurately. For SDN also machine learning can play part by predicting attacks in an early stage and prevent data lose.

Block Diagram of SDN



What problem to be solved?

- SDN, one of the most populist networking model is not risk free.
- It is used by many big companies and government sectors.
- Hackers try to take control of it in order to get confidential information of big companies or even government sectors.
- As a result, many companies or even people of a certain countries information can be gained and be used to hamper a nation’s security.

How the problem will be solved

- As discussed earlier, SDN is not risk free.
- Many information can get into wrong hand without even the owner realizing it.
- As it is used by big companies and even government sectors, an entire nation's security will be at stake.
- However if the attack can be predicted early than this data lose can be prevented.
- Machine learning algorithms like Random Forest, Decision Tree can help to predict an attack. Even if not early at least it can predict in an early stage. As a result many data lose can be prevented.

What was done?

- We have implemented four machine learning models (Random Forest, Decision Free, Gradient Boosting and Ada Boosting) into our database collected from Kaggle.
- We got 99.38% accuracy for Random Forest and 99.24% accuracy for Decision Tree, 99.72% for Gradient Boosting and 99.34% accuracy for Ada Boosting as shown in the above slides.
- It concludes Gradient Boosting is the most accurate.

Results (Random Forest's Accuracy)

```
[ ]  Rclf.fit(xtrain,ytrain)

     RandomForestClassifier()

[ ]  Rclf.score(xtest,ytest)

     0.9938112177769115
```

Results (Decision Tree's Accuracy)

```
[ ]  Clf.fit(xtrain,ytrain)
```

```
DecisionTreeClassifier()
```

```
[ ]  Clf.score(xtest,ytest)
```

```
0.9924873706609489
```

Results (Gradient Boosting's Accuracy)

```
GradientBoostingClassifier()
```

```
[ ]  grad_model.score(xtest,ytest)
```

```
0.9972035794183445
```

Results (Ada Boosting's Accuracy)

```
AdaBoostClassifier()
```

```
ada.score(xtest,ytest)
```

```
0.9934428758774975
```

Results (Random Forest's F1 score)

	precision	recall	f1-score	support
Attack	1.00	1.00	1.00	3649
Normal	1.00	1.00	1.00	3639
accuracy			1.00	7288
macro avg	1.00	1.00	1.00	7288
weighted avg	1.00	1.00	1.00	7288

Results (Decision Tree's F1 score)

	precision	recall	f1-score	support
Attack	1.00	1.00	1.00	3649
Normal	1.00	1.00	1.00	3639
accuracy			1.00	7288
macro avg	1.00	1.00	1.00	7288
weighted avg	1.00	1.00	1.00	7288

Results (Gradient Boosting's F1 score)

	precision	recall	f1-score	support
Attack	1.00	1.00	1.00	3649
Normal	1.00	1.00	1.00	3639
accuracy			1.00	7288
macro avg	1.00	1.00	1.00	7288
weighted avg	1.00	1.00	1.00	7288

Results (Ada Boosting's F1 score)

	precision	recall	f1-score	support
Attack	1.00	1.00	1.00	3649
Normal	1.00	1.00	1.00	3639
accuracy			1.00	7288
macro avg	1.00	1.00	1.00	7288
weighted avg	1.00	1.00	1.00	7288

Comparison

- Four machine learning models was implemented. Those were: Random Forest, Decision Free, Gradient Boosting and Ada Boosting.
- 99.72% for Gradient Boosting, 99.38% accuracy for Random Forest, 99.34% accuracy for Ada Boosting and 99.24% accuracy for Decision Tree.
- It concludes Gradient Boosting is the most accurate. The second best model was Random Forest, the third best model was Ada Boosting and the last one of the list was Decision Tree.

Tools Used For The Project

- Google Collab
- PyCharm
- Jupyter NoteBook

Funding

- There won’t be any funding required as it is a software based project.

Novelty

- The project has better accuracy than the existing projects available for SDN intrusion detection.
- We have also used two machine learning models Gradient bosting and Ada boosting which were not used earlier for intrusion detection.

Conclusion

- SDN which is the short form of software defined network is one of the most populist model.
- As it is becoming popular, unethical people are giving their eye on it.
- Hackers plan to take control of the model and access many information.
- As a result many people or even a nation’s security is at stake.
- Our system will take help of various machine learning algorithm to predict the attack before it takes place or at earliest stage
- As a result many data lose can be prevented.

Thank you