BH-PCMLAI: University of California, Berkeley

Module 24: Capstone Project Presentation

BGP Routing Anomaly Detection System

Shaji Ravindra Nathan

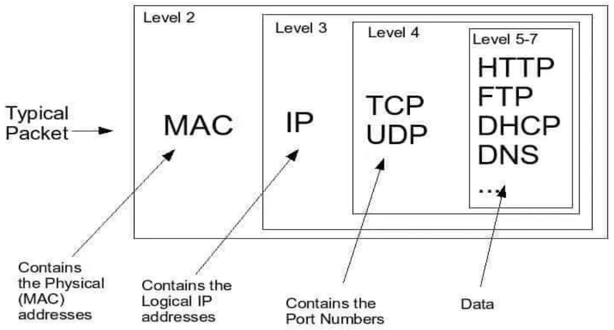
Section B

11/26/2022

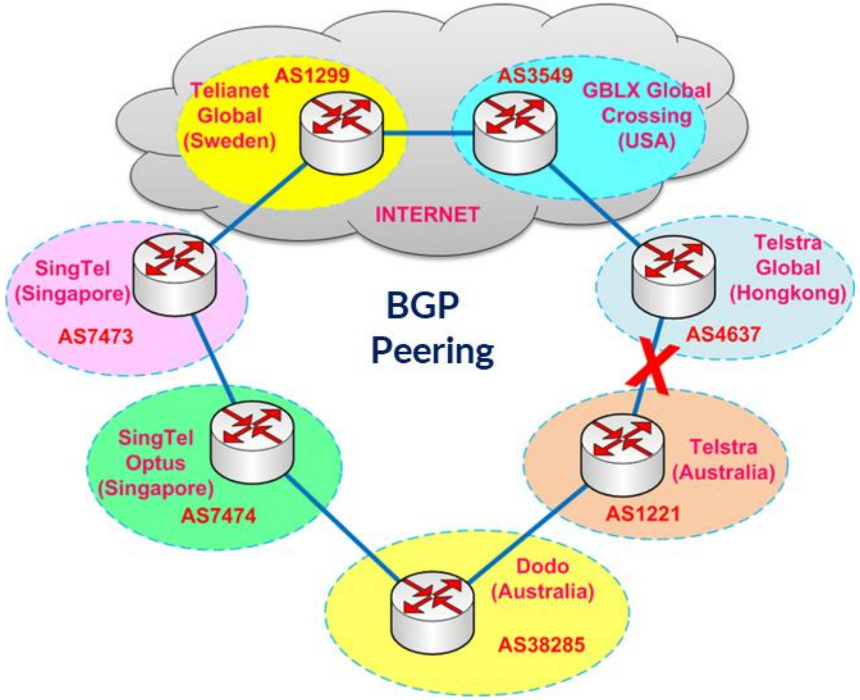
INDEX

BGP Routing Explained

Border Gateway Protocol — the "post office of the internet"
Responsible for sending information in the form of "packets."

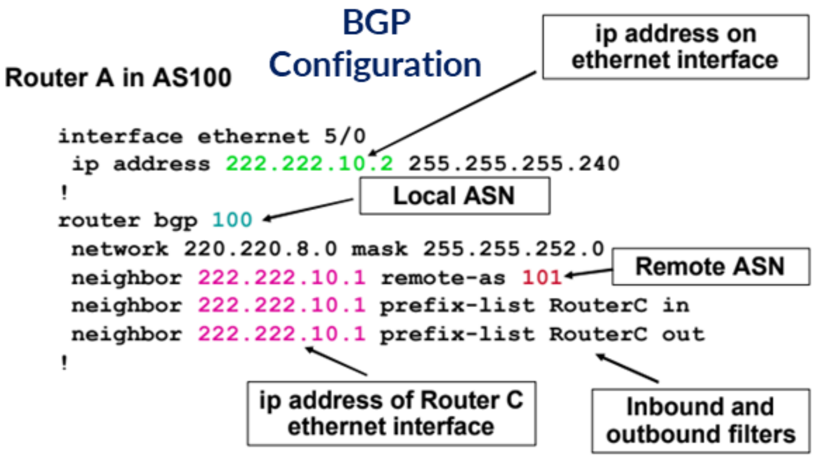


- **In simple terms, internet traffic consists in hundreds of millions of routers and switches communicating with another using digital signals.**
- **These signals are arranged in packets, which are sequences of 1s and 0s of a designated length.**



Border Gateway Protocol Core Functions:

- Enabling coordination between the over 70,000 different networks that interconnect into the single global communication infrastructure that we call the internet.
- Standardized way to exchange routing and reachability information among networks which make up the internet a.k.a autonomous systems (AS) on the Internet.
  For example: ATT has an ASN # 7018
- Make routing decisions based on network paths, network policies, or rule-sets configured by a network administrator.
- Guide Internet Protocol packets from end user to the final destination across the internet



Problem Statement: BGP related Malicious Activity in the news

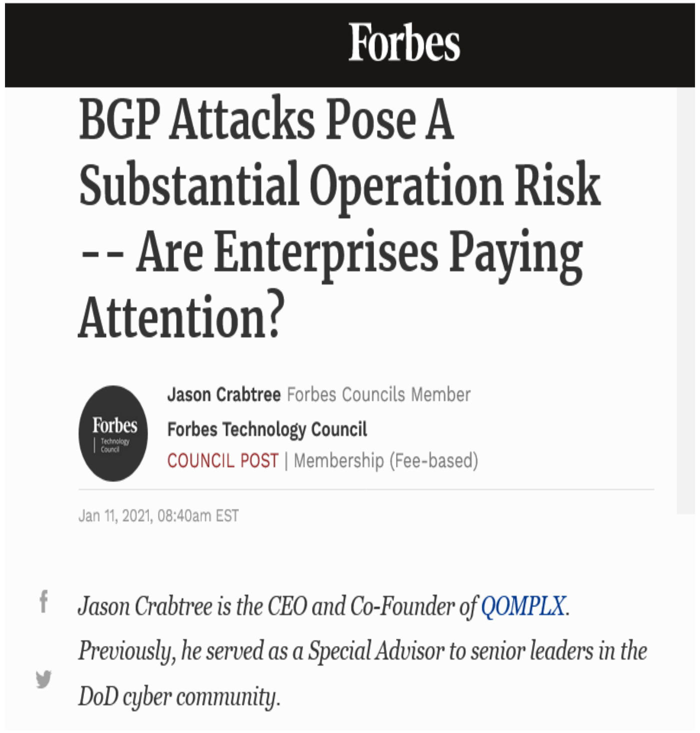
Edge Articles | 8 MIN READ | ARTICLE

101: Why BGP Hijacking Just Won't Die

A look at the dangers of attacks on the Internet's Border Gateway Protocol and what ISPs and enterprises can do about them.

Seth Rosenblatt
Contributing Writer

July 28, 2021



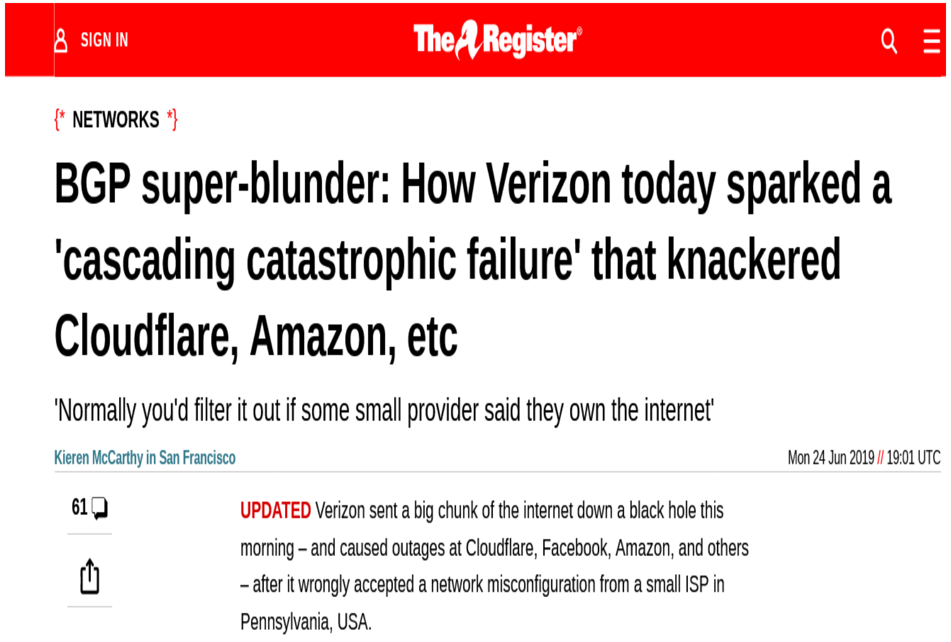
Forbes

BGP Attacks Pose A Substantial Operation Risk -- Are Enterprises Paying Attention?

Jason Crabtree Forbes Councils Member
Forbes Technology Council
COUNCIL POST | Membership (Fee-based)

Jan 11, 2021, 08:40am EST

Jason Crabtree is the CEO and Co-Founder of QOMPLX. Previously, he served as a Special Advisor to senior leaders in the DoD cyber community.



SIGN IN

The Register

{* NETWORKS *}

BGP super-blunder: How Verizon today sparked a 'cascading catastrophic failure' that knackered Cloudflare, Amazon, etc

'Normally you'd filter it out if some small provider said they own the internet'

Kieren McCarthy in San Francisco

Mon 24 Jun 2019 // 19:01 UTC

61

UPDATED Verizon sent a big chunk of the internet down a black hole this morning – and caused outages at Cloudflare, Facebook, Amazon, and others – after it wrongly accepted a network misconfiguration from a small ISP in Pennsylvania, USA.

Financial Impact Example:

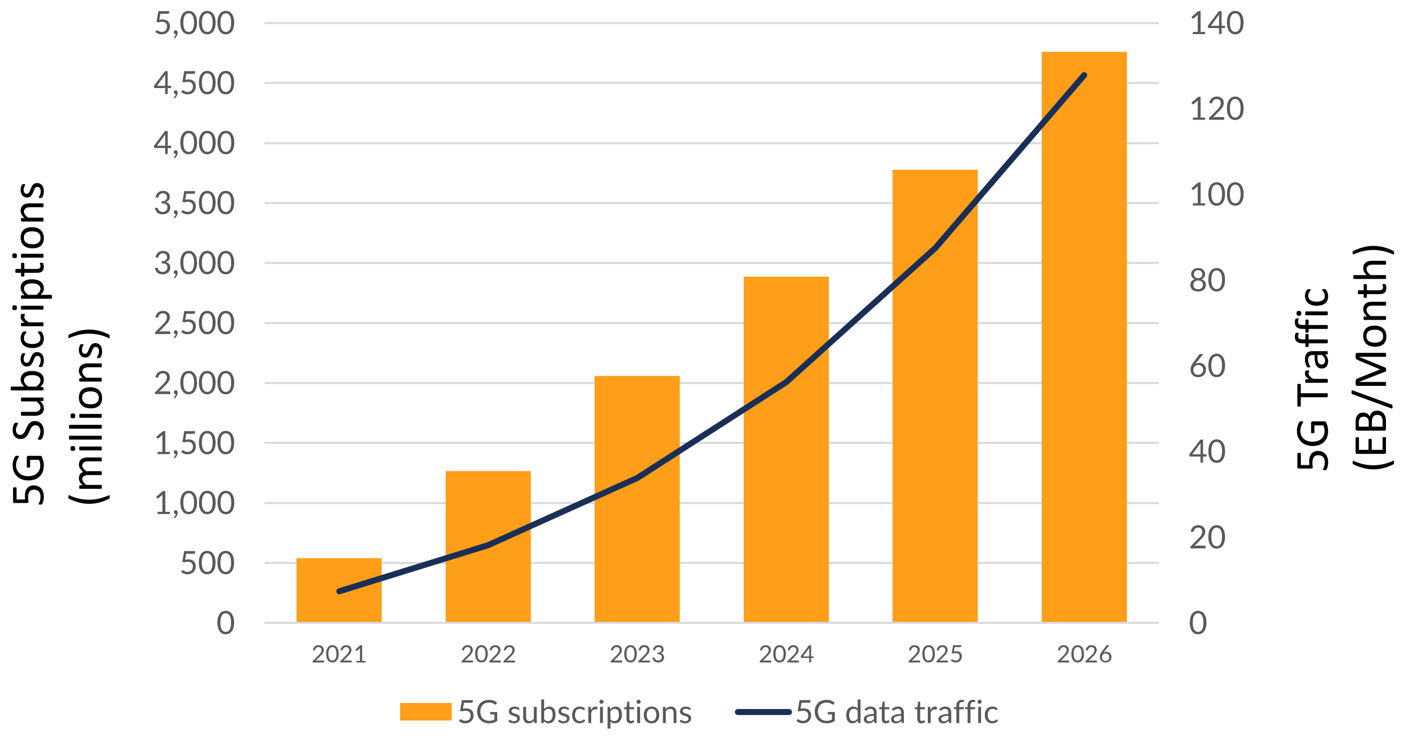
- **On Monday, October 4 2021, a BGP outage cost Facebook roughly $60 million in revenues over its more than six-hour period.**
- **Facebook shares fell 4.9 percent on the day, which translated into more than $47 billion in lost market cap.**
- Reference: https://www.datacenterdynamics.com/en/opinions/too-big-to-fail-facebooks-global-outage/
- Live Outage Map: This map shows outages at any given time on the internet
- https://www.thousandeyes.com/outages/?utm_source=Blog&utm_medium=Textlink&utm_campaign=Most-Disruptive-Internet-Outages-2020

Too much data and no time

Internet traffic data is logged, however, the scale is far too great to actively monitor and react:

- 100Gbps allows a single pathway to present one valid 64 octet IP packet every 5 nanoseconds
- It is a near impossibility to inspect the veracity of all of it without introducing unacceptable latency
- As internet usage proliferates around the globe, the vulnerabilities will only increase

Projected Global 5G Subscriptions and Data Traffic Forecasts



So how do we distinguish malicious activity from legitimate usage?

Machine Learning Based Solution:

BGP Anomaly Detector
for
Routing Security

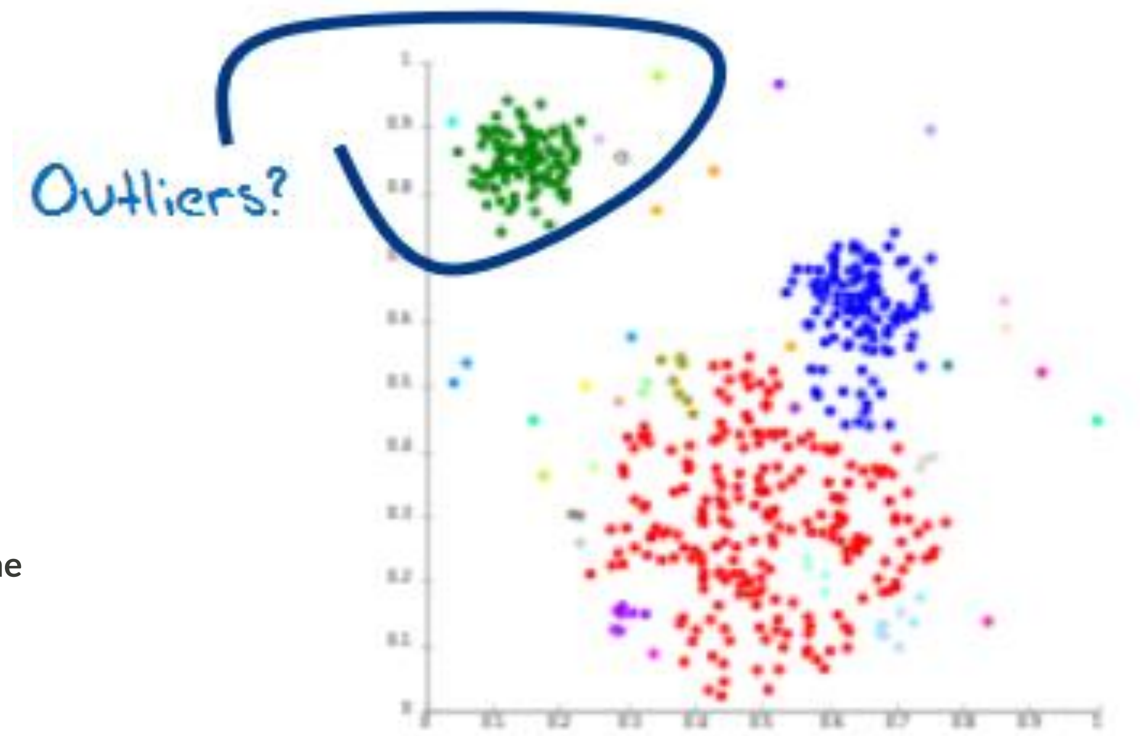
Machine Learning (ML) for Anomaly Detection

What it does :

- Detects BGP routing security breaches, anomalies

How it is done:

- Initial goal is to use a supervised machine learning model
- That does Real time Classification of BGP Updates
- Separates routing traffic into two classes
  - Normal
  - Anomalous Traffic
- Labeled training samples used to learn a classification hyperplane
- Machine Model developed from real time BGP updates



How it is deployed:

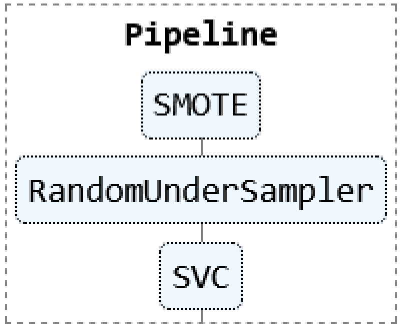
- A standalone classification software package that works with ACLs and BGP module in OCNOS (Open Compute Network Operating System)
- This software can be pushed to interact with OCNOS control plane via a docker container
- Can be Sold as an upsell application on top of OCNOS

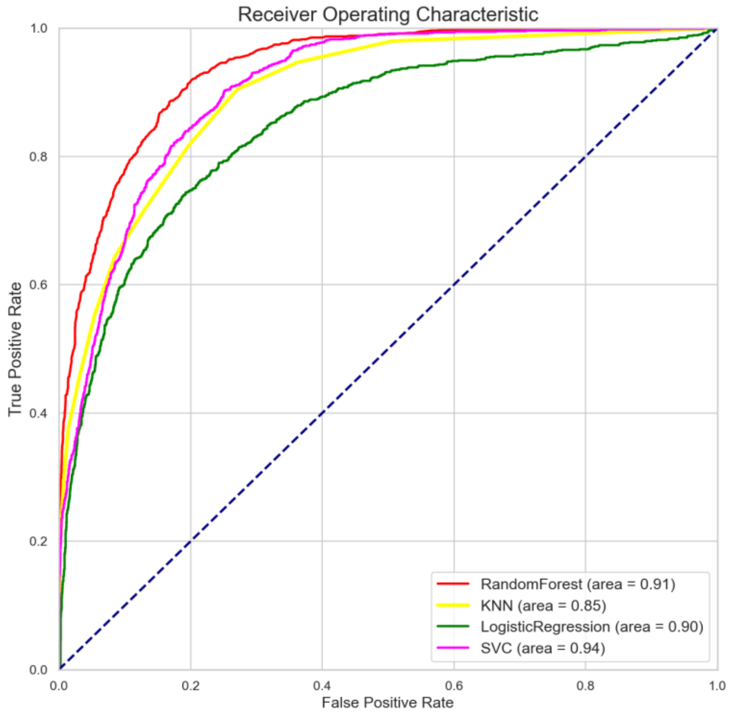
Methodology Used

ipinfusion

Number of withdrawals	Number of announced NLRI prefixes	Number of withdrawn NLRI prefixes	Average AS-path length	Maximum AS-path length	...	Maximum AS-path length11	Maximum AS-path length12	Maximum AS-path length13	Maximum AS-path length14	Maximum AS-path length15	Number of Interior Gateway Protocol (IGP) packets	Number of Exterior Gateway Protocol (EGP) packets	Number of incomplete packets	Packet size (B)	Label
7	1761	358	6	16	...	0	0	0	0	1	449	0	40	302	1
6	97	19	6	10	...	0	0	0	0	0	43	0	14	228	1
8	802	52	6	15	...	0	0	0	1	0	256	0	21	266	1
7	206	198	6	10	...	0	0	0	0	0	41	1	12	354	1
6	266	35	7	22	...	0	0	0	0	0	76	0	14	265	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
5	202	31	5	10	...	0	0	0	0	0	40	0	1	303	-1
3	286	5	6	7	...	0	0	0	0	0	31	0	0	348	-1
6	96	42	6	8	...	0	0	0	0	0	41	0	3	247	-1
4	170	24	6	9	...	0	0	0	0	0	58	0	7	255	-1
6	89	24	6	9	...	0	0	0	0	0	54	0	3	223	-1

- A dataset with 42 features was labelled for Nimda, Codred and WannaCry
- Multiple models were built using Supervised Learning
  - Logistic Regression
  - RandomForest
  - KNN
  - Support Vector Machine
- SVM was selected as it had the best performance

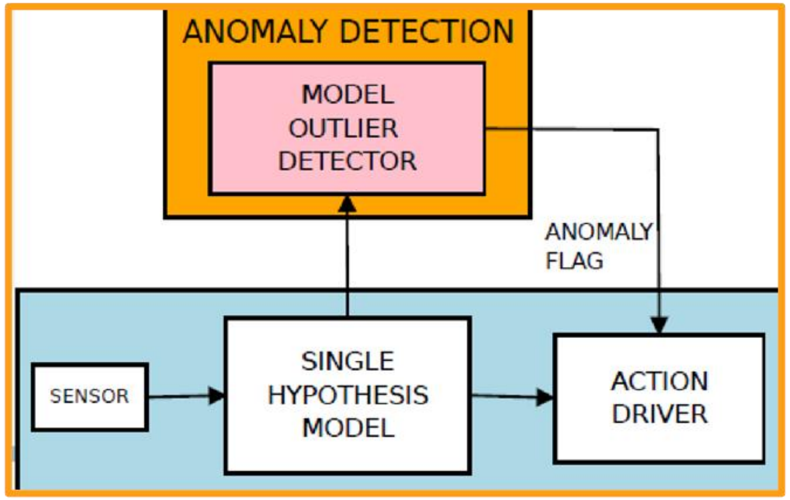




Support Vector Machine Explained

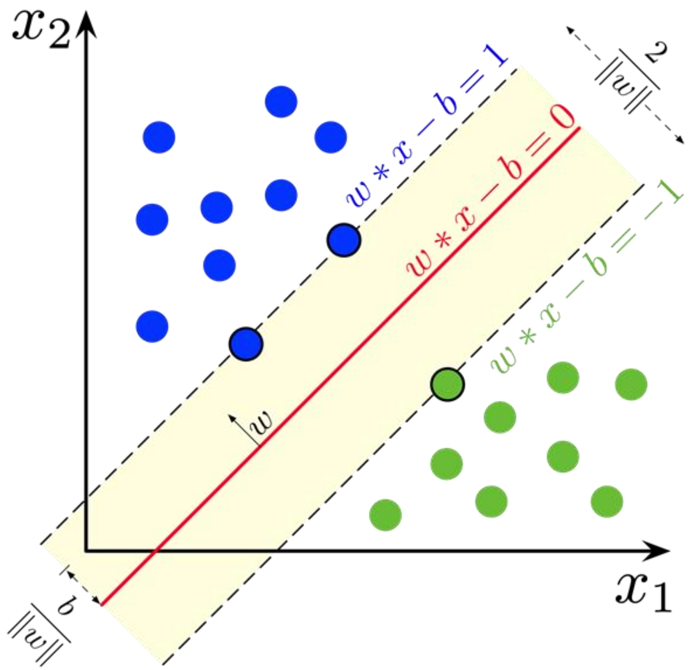
- Anomaly Detector High Level Functional Blocks

- The idea is to come up with a two class classifier that classifies BGP updates with
  - 1 indicating an anomaly and -1 indicating normal update



H_0 ©μ } 657 events

H_A ©μ ®657 events



We are given a training dataset of n points of the form

$$(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n),$$

Any hyperplane can be written as the set of points $\mathbf{x}$ satisfying

$$\mathbf{w}^{\mathsf{T}}\mathbf{x} - b = 0,$$

Support Vector Machine Explained

- SVM Kernel Trick (Radial Basis function)
- Transforming data to make it linearly separable

A symmetric function $K(x, y)$ can be expressed as an inner product

$$K(x, y) = \langle \phi(x), \phi(y) \rangle$$

for some ϕ if and only if $K(x, y)$ is positive semidefinite, i.e.

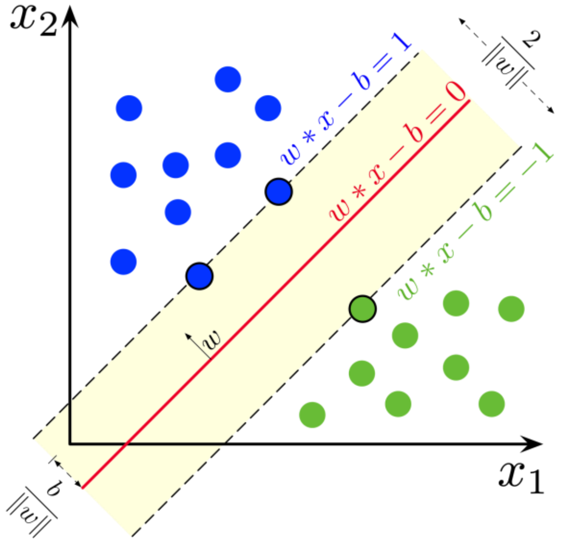
$$\int K(x, y) g(x) g(y) dx dy \geq 0 \qquad \forall g$$

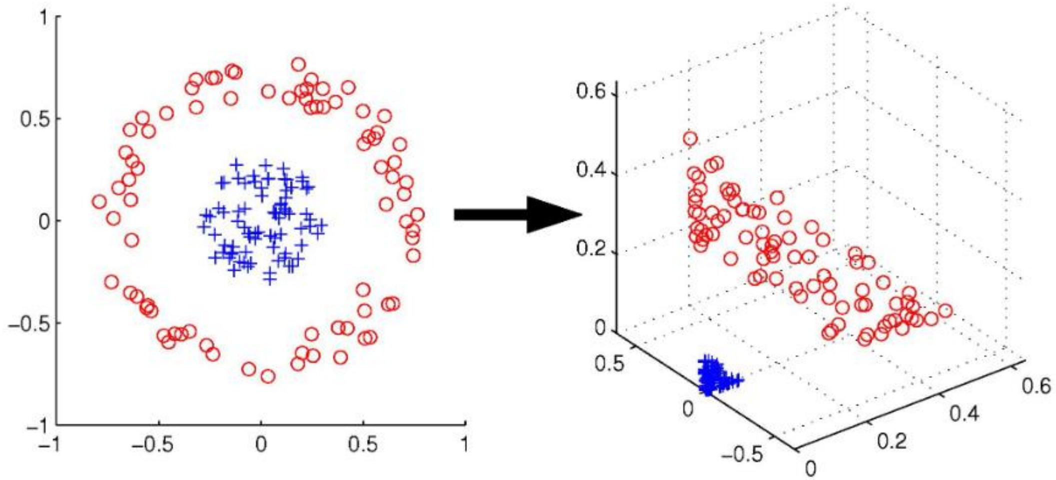
or, equivalently:

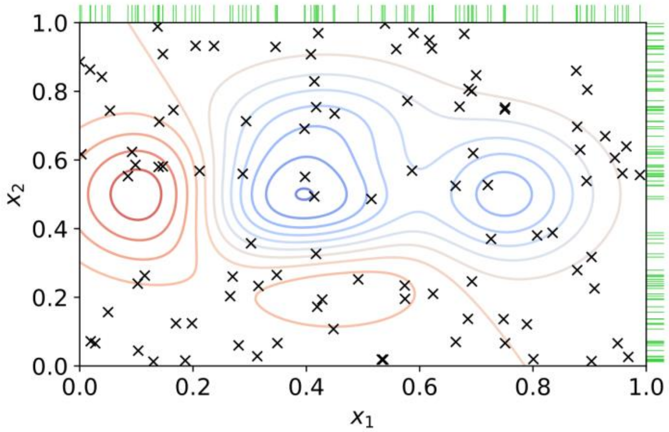
$$\begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) & \cdots \\ K(x_2, x_1) & \ddots & \\ \vdots & & \end{bmatrix} \text{ is psd for any collection } \{x_1 \ldots x_n\}$$

Therefore you can either explicitly map the data with a ϕ and take the dot product, or you can take any kernel and use it right away, without knowing nor caring what ϕ looks like. For example:

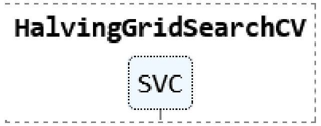
- Gaussian Kernel: $K(x, y) = e^{\frac{1}{2}||x-y||^2}$
- Spectrum Kernel: count the number of substrings in common. It is a kernel since it is a dot product between vectors of indicators of all the substrings.







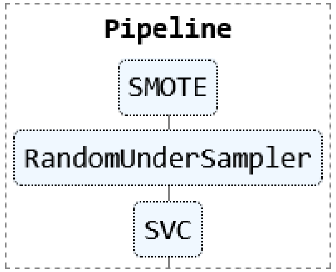
Methodology Used

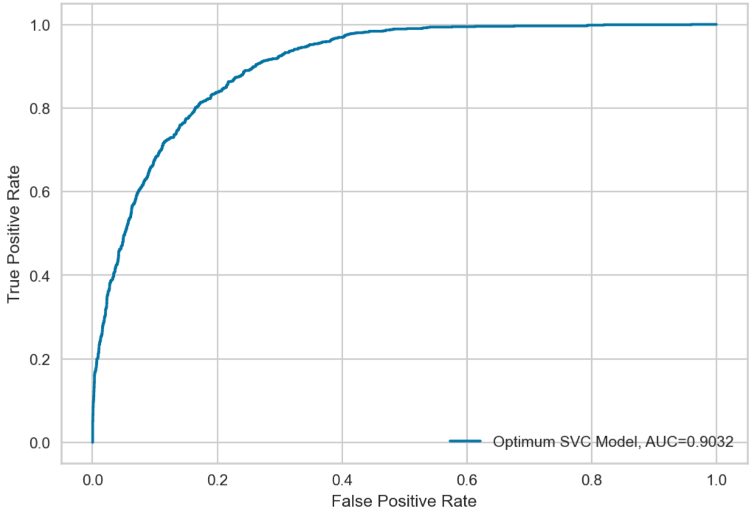


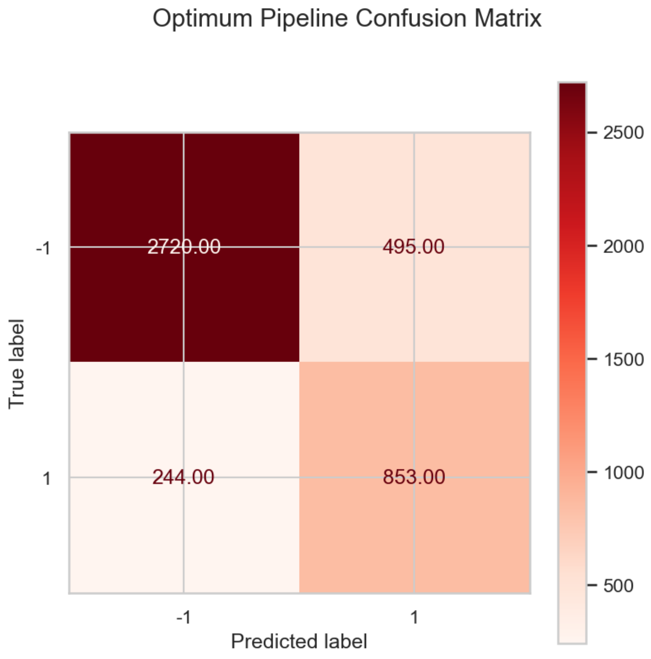

```
print("Best cross-validation accuracy: {:.3f}".format(halving_grid_svc.best_score_))
print("Test set score: {:.3f}".format(halving_grid_svc.score(X_test, y_test)))
print("Best parameters: {}".format(halving_grid_svc.best_params_))
```

```
Best cross-validation accuracy: 0.852
Test set score: 0.850
Best parameters: {'C': 1, 'gamma': 0.0001, 'kernel': 'rbf'}
```

- GridSearch was used to do the best hyperparameter selection
- SMOTE library was used to balance the classes
- The parameters were used to build the final model
- This new model was used for predicting new BGP updates








```
#setup a pipeline based on GridSearch suggested results
optimum_svc_model = SVC(C= 1, degree=2, gamma='scale', probability=True, verbose=True,kernel='rbf')
optimum_pipeline = Pipeline([('over', SMOTE(random_state=42)),
                             ('under', RandomUnderSampler(random_state=42)),
                             ('clf', optimum_svc_model)])
optimum_pipeline.fit(X_train, y_train)
```

Result Summary

- The resulting model is able to classify new BGP routing update traffic from a cellular gateway like so:
  - -1 for normal traffic
  - 1 for anomalous traffic
- Example below

New BGP Update

	Hour and Minutes	Hour	Minutes	Seconds	Number of announcements	Number of withdrawals	Number of announced NLRI prefixes	Number of withdrawn NLRI prefixes	Average AS-path length	Maximum AS-path length	...	Maximum AS-path length11	Maximum AS-path length12	Maximum AS-path length13	Maximum AS-path length14
17237	2350	23	50	6	44	5	91	19	7	14	...	0	0	1	0
17238	2351	23	51	1	42	5	63	21	6	14	...	0	0	1	0
17239	2352	23	52	4	29	4	40	8	6	10	...	0	0	0	0
17240	2353	23	53	1	59	4	99	15	7	14	...	0	0	1	0
17241	2354	23	54	14	41	4	56	11	6	12	...	1	0	0	0
17242	2355	23	55	12	41	5	202	31	5	10	...	0	0	0	0
17243	2356	23	56	5	31	3	286	5	6	7	...	0	0	0	0
17244	2357	23	57	0	44	6	96	42	6	8	...	0	0	0	0
17245	2358	23	58	4	65	4	170	24	6	9	...	0	0	0	0
17246	2359	23	59	14	57	6	89	24	6	9	...	0	0	0	0

10 rows × 42 columns

Classifier in action

In [105]:
```
newbgpupdate=df[17237:]
newbgpupdate= newbgpupdate.drop("Label", axis=1)
```

In [106]:
```
Grid_Optimized_Prediction(newbgpupdate)
```

```
The predicted bgp update status is: $ [-1 -1 -1 -1 -1 -1 -1 -1  1 -1]

Here is how to interpret the results:

An outcome of  '1' indicates that there is an Anomaly in this BGP Update

An outcome of '-1' indicates that the BGP Update is normal
```

Results Summary: Actionable Insights

1. The objective of this capstone project was to come up with an optimum classification model to predict whether a bgp update message can be classified as anomalous or normal based on the update message attributes.
2. We analyzed 37 numerical variables which capture the various attributed of a typical BGP protocol update to build the model.
3. Exploratory data analysis showed absence of null values in the dataset, and the data is imbalanced, where "1" anomalous message is the majority class.
4. Univariate analysis revealed that Average_AS_Path length and Number of Implicit withdrawals does not help very much when it comes to predicting the target variable. Some numerical features tend to predict the target variable much better (for example: ['Maximum AS-path length', 'Number of duplicate announcements', 'Maximum edit distance', 'Number of Exterior Gateway Protocol (EGP) packets'] etc.)
5. Dataset preprocessing of numerical data was done using standscaler and MinMax scaler
6. Basic models were built using K Nearest Neighbor, Logistic Regression, Decision Trees, and Support Vector Machines.
7. The most important features in predicting whether a BGP update is anomalous based on the Support Vector model was ['Hour and Minutes', 'Hour', 'Minutes', 'Seconds',

```
'Number of announcements', 'Number of withdrawals',
'Number of announced NLRI prefixes',
'Number of withdrawn NLRI prefixes', 'Average AS-path length',
'Maximum AS-path length', 'Average unique AS-path length',
'Number of duplicate announcements',
'Number of duplicate withdrawals',
'Number of implicit withdrawals']
```

8. GridSearch and Halving GridSearch was used to find the best parameters. The best parameters derived from Gridsearch were as follows: {'C': 1, 'gamma': 0.0001, 'kernel': 'rbf'}. SMOTE library was used to remedy the class imbalance.
9. Radial Basis Kernel function was used to find a non-linear classifier, C was set to 1 to control error based on the RIPE dataset, to minimize the cost function. Gamma was set to a low value to figure out the optimum curvature in the decision boundary.
10. Support Vector model gave the best performance with Halving GridSearchCV and the best test AUC was 0.86 which was similar to the results reported by researchers at CAIDA, RIPE, and Simon Fraser University in previous research reports.

Next Steps

Next Steps in the Classifier Model Enhancement:

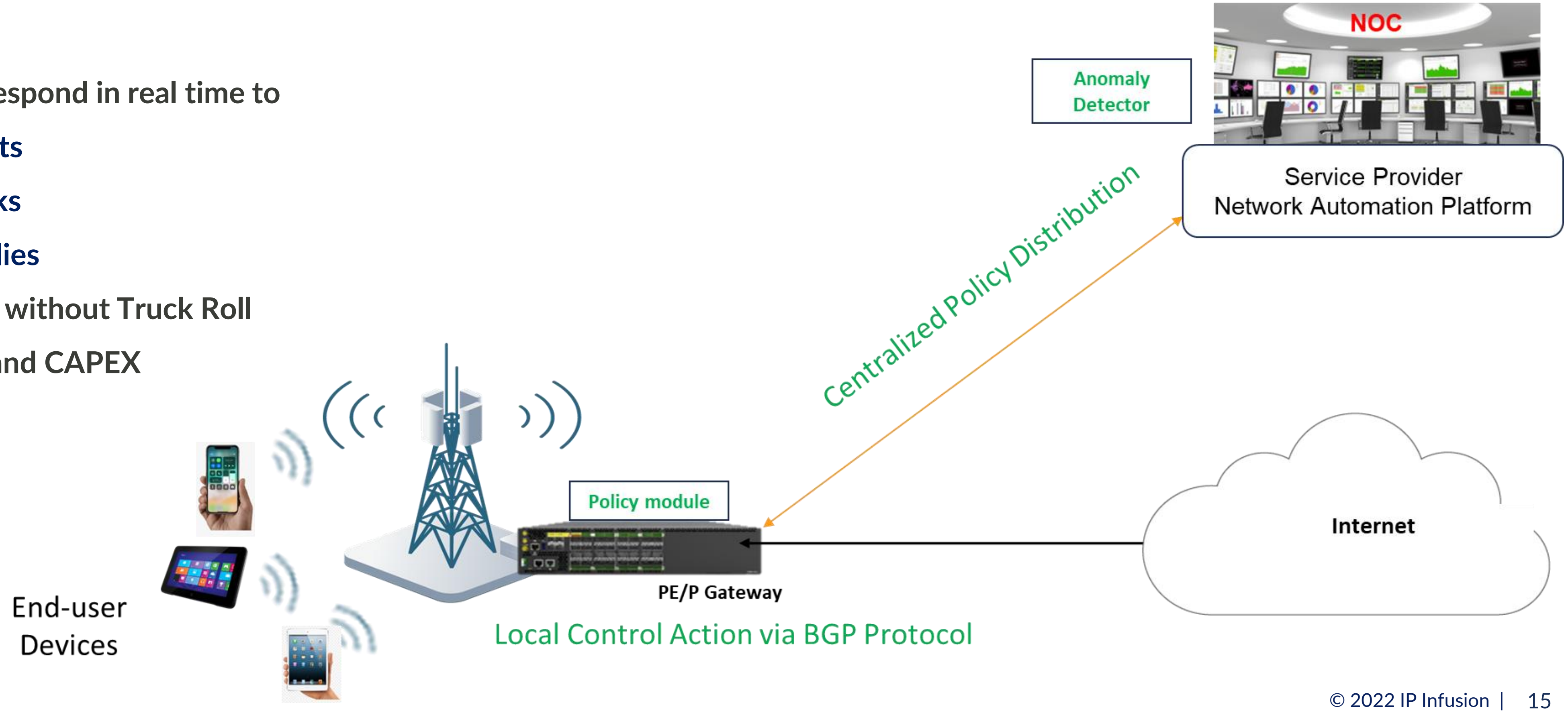
1. Fine Tuning the model based on domain knowledge and feature importance results
2. We could reduce the dimensions/features further to tune the model.
3. Support Vector Model training even with HalvingGridSearch was very slow.
4. There are some new libraries like T-POT https://epistasislab.github.io/tpot/using/ that use genetic algorithms for hyperparameter tuning to derive the best pipeline for this classification problem. For best feature selection we could use a library like YellowBrick https://www.scikit-yb.org/en/latest/api/model_selection/importances.html
5. Ensemble models, XGBoost could be used as next step in improving the performance of the current SVM model
6. Neural Networks/Autoencoders and LSTM based model seems to be well suited for this class of problems.

Network Deployment : Cellular Backhaul Gateway Protection

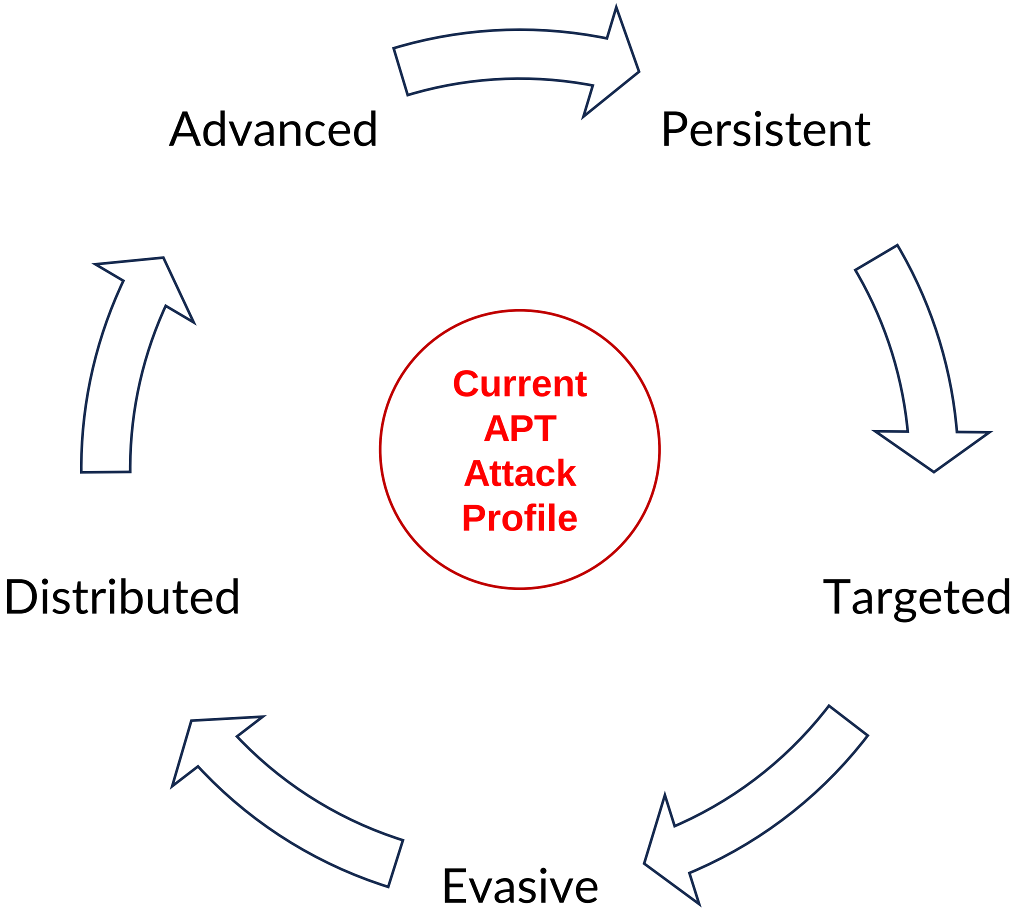
- Anomaly Detector is deployed in the Cellular Data Center
- BGP updates are provided to the AD pipeline by a BGP agent
- BGP agent does the ETL on the packet header
- Provides Realtime detection of BGP Anomalies/Malicious attacks
- Cell Gateway autonomously blocks/reroutes anomalous traffic

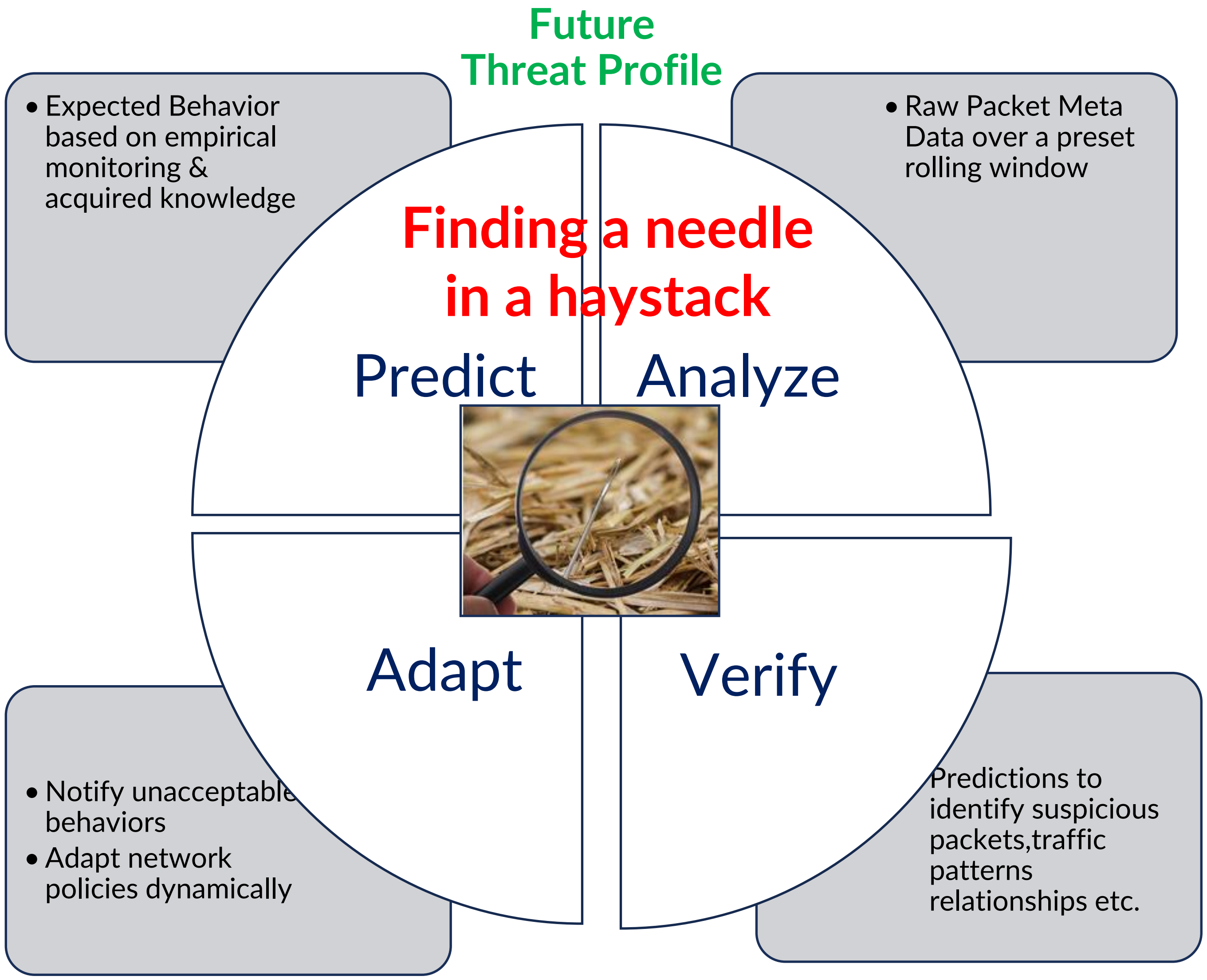
Business Benefit:

- Enables Telco Carriers to respond in real time to
  - Emerging Threats
  - Zero Day Attacks
  - Routing Anomalies
- Prevents Loss of Service without Truck Roll
- Lowers Security OPEX and CAPEX



The Business Benefit:





Future Use Cases in Telecom using AI/ML Techniques

System Monitoring	Managed Services	Intelligent Networks
Anomaly Detection	Ticket Classification	Self-Healing
Root Cause Identification	Churn Prediction	Dynamic Optimization
Predictive Maintenance	SLA Assurance	Automated Network Design

Current Project

Thanks !

contact: shaji.nathan@ipinfusion.com
phone: 408.400.1503