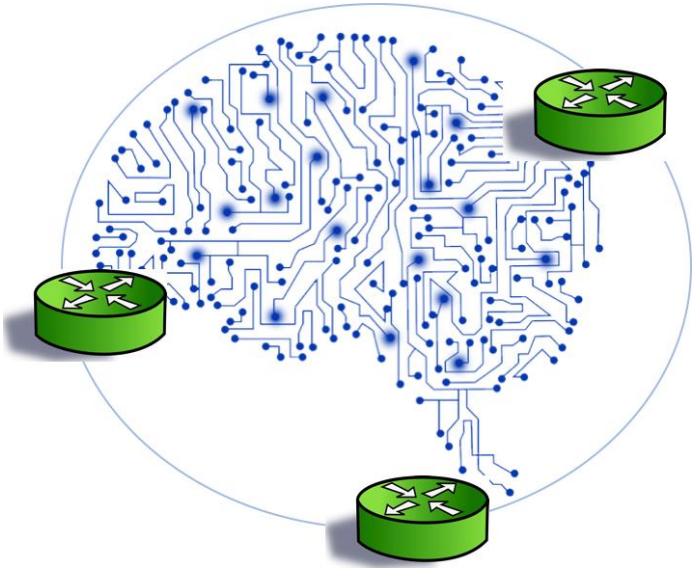


Machine Learning for Network Operations

Jun 2022
Introductory and Advanced Topics on P4 Programmable Data Plane Switches Workshop

Dr. Mariam Kiran, Prof Jorge Crichigno (UofSC)
<mkiran@es.net>
Scientific Networking Division
Lawrence Berkeley National Laboratory

BERKELEY LAB

U.S. DEPARTMENT OF ENERGY
Office of Science

Outline

Understand what is Machine Learning (ML) and how can we use it in our experiments

Introduction to AI and ML

Summary of techniques and libraries used

Example of ML tasks

Measuring accuracy of ML models

Example of ML in networks

Highlight of Deep learning and Network explorations

What is AI (Artificial Intelligence)?

Basic Definition of AI: “Machines start exhibiting intelligent behavior”

- Alan Turing‘s paper “Can Machines Think!”
  - Posed this question in 1950
- Turing Test:
  - Cannot determine if you are talking to human or machine

AI versus ML versus DL

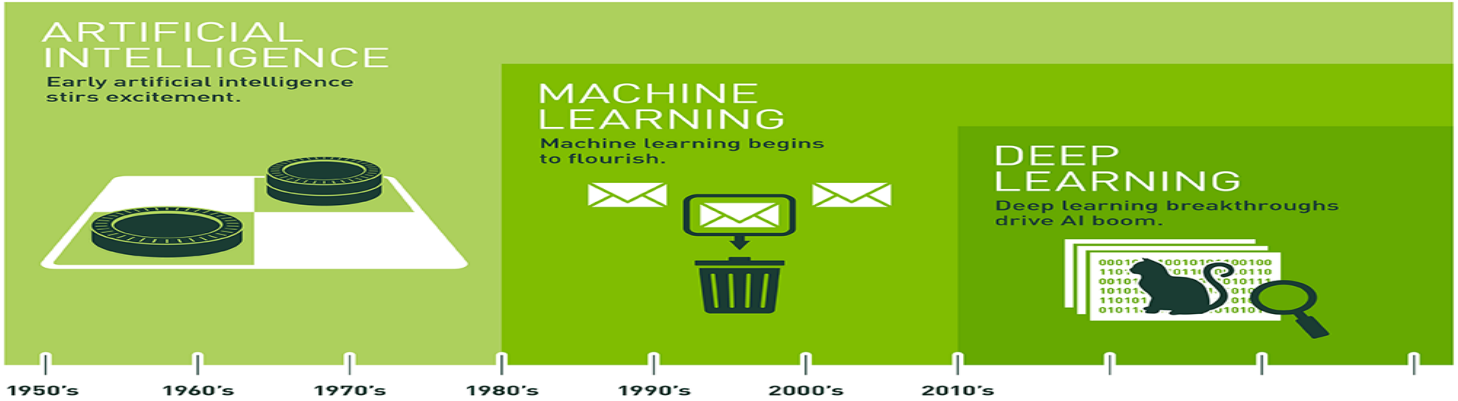
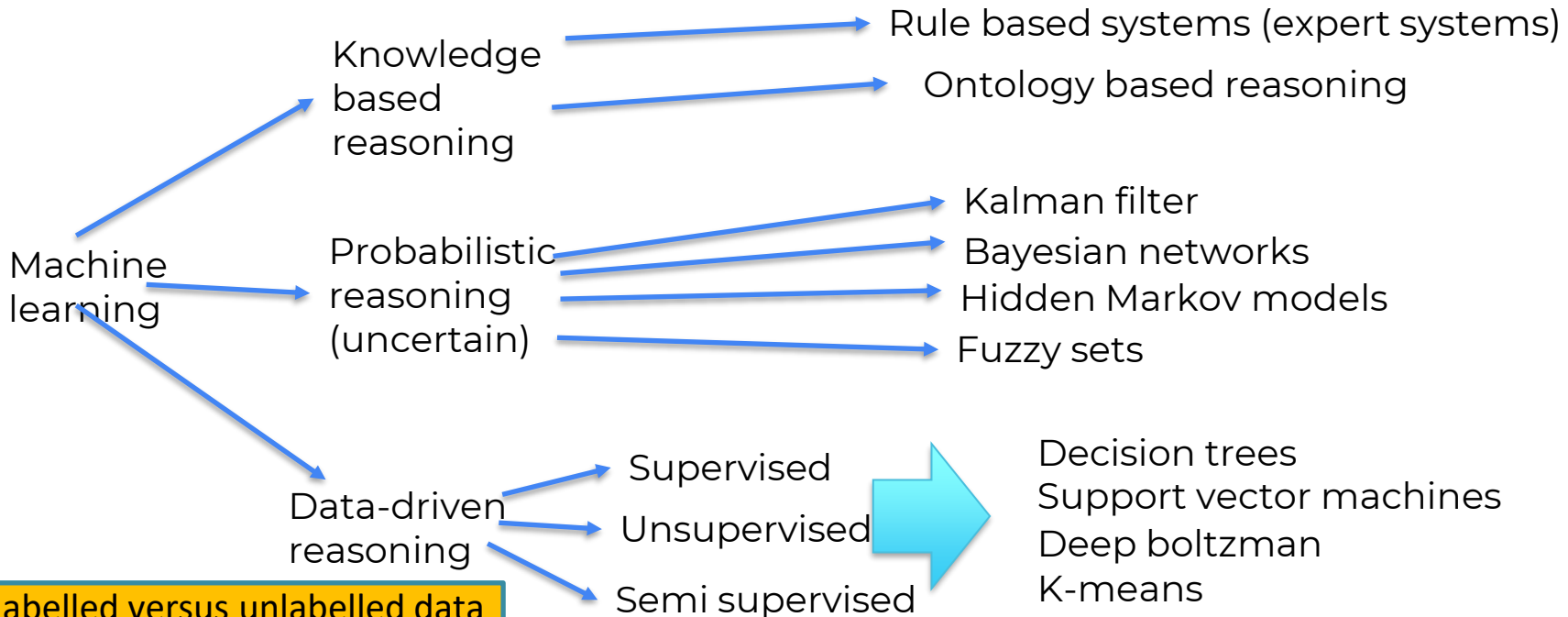


Image from Nvidia Blog

- Machine learning is an approach to achieve AI – spam filters, HR
- Deep learning is one of the techniques for ML:
- Recent advances due to GPU and HPC processing (previously very slow, too much data, need training to work)
- Mainly for image and speech recognition – commercial apps

Machine Learning Categories

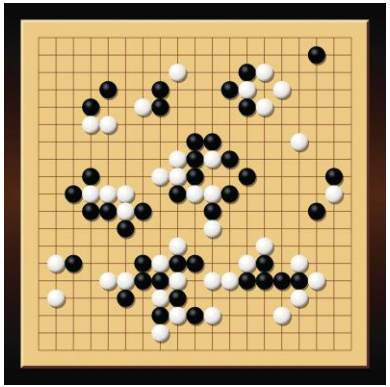


- Labelled versus unlabelled data sets
- Depends on the data available
- Differs based on problems

Why Deep Learning (DL)?

Trained to recognize cats, "object identification"

Play games, "Best strategies for winning game"



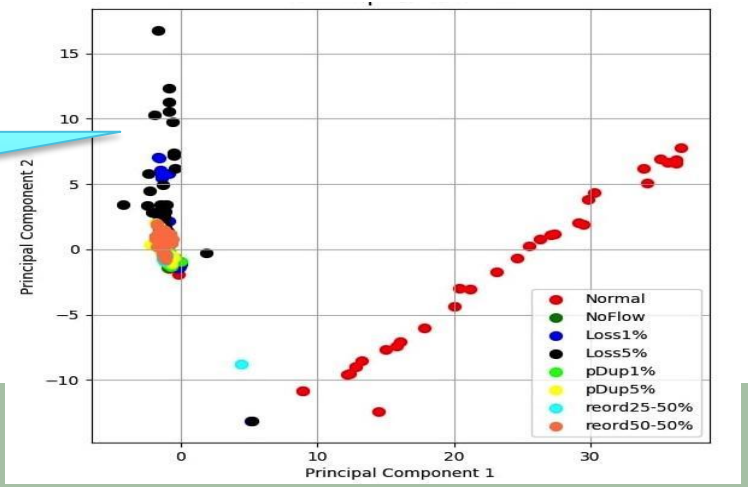
Deep learning (neural networks - NN) introduces 'data-driven learning' to build bespoke solutions

Supervised Learning

When Labelled Data is Available

- Learning to recognize the labels
- Classification, object detection, anomaly detection
- Works very well if we can identify clear class boundaries

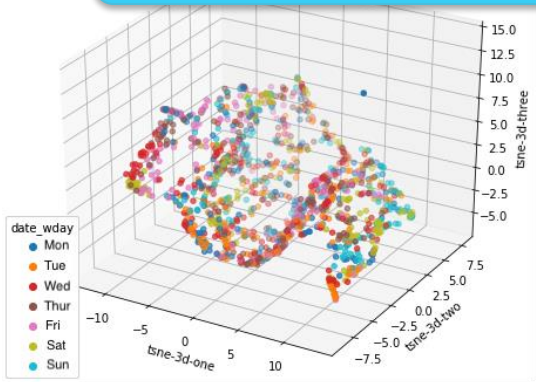
Good versus bad flows



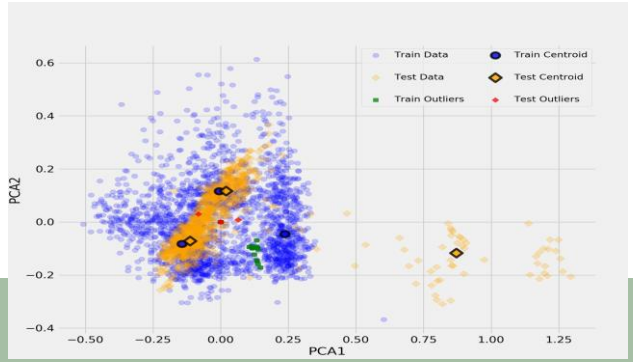
Plot of anomalous TCP traffic

Unsupervised Learning

Visualizing flows into clusters



Plot of TCP flow transfers



When NO labelled data in Available

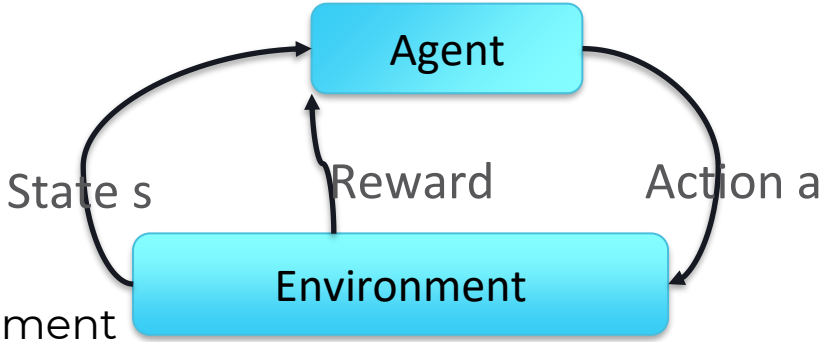
- Learn underlying rules in the data
- Clustering, feature identification, recognize anomalies in test data

- Works well when working with domain scientists after clusters are recognized

Reinforcement Learning

When NO data is available

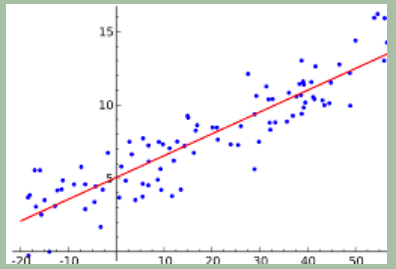
- Learn via trial and error
- Interactions with the environment

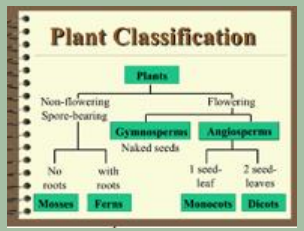


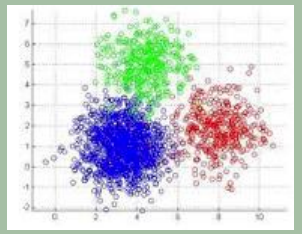
- Agent views the state of the environment and chooses an action
- Each action, agent gets a reward
- Over time, agent learns optimal actions that give best rewards and what to do next

Common ML Tasks

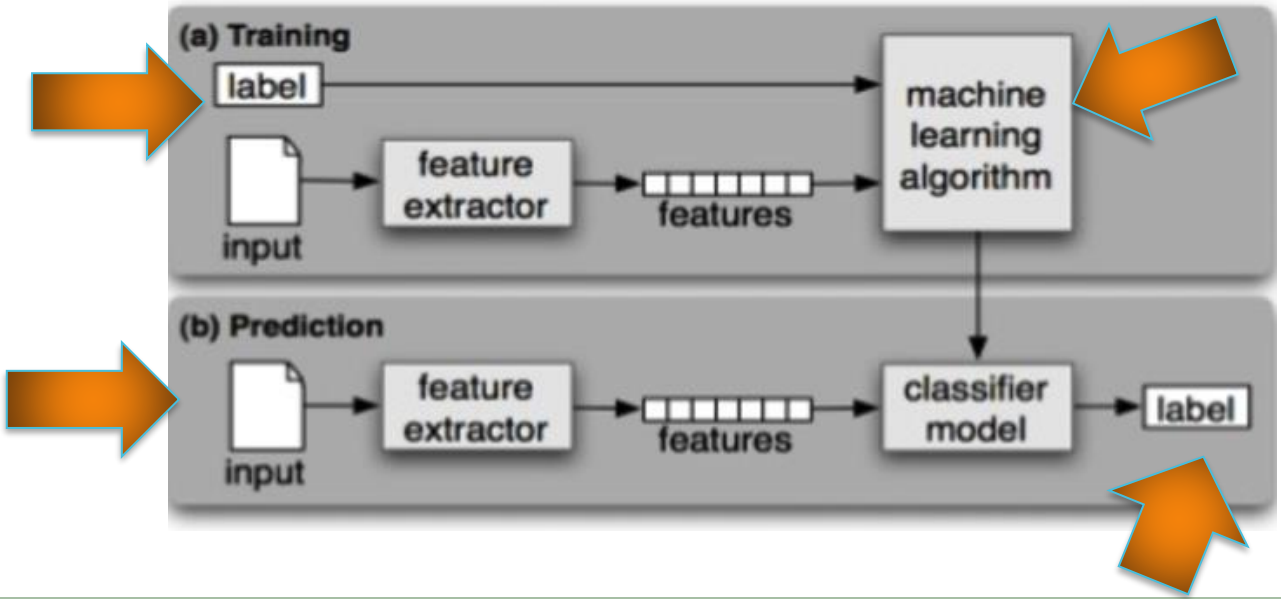
- Regression
  - Predict value in the future
  - Gaussian, poisson, linear graphs....
- Classification
  - data into classes/categories (supervised)
  - Ranking, neural networks, decision tree...
- Clustering
  - partition data into clusters (unsupervised)
  - K-means, neural networks, nearest-neighbor...

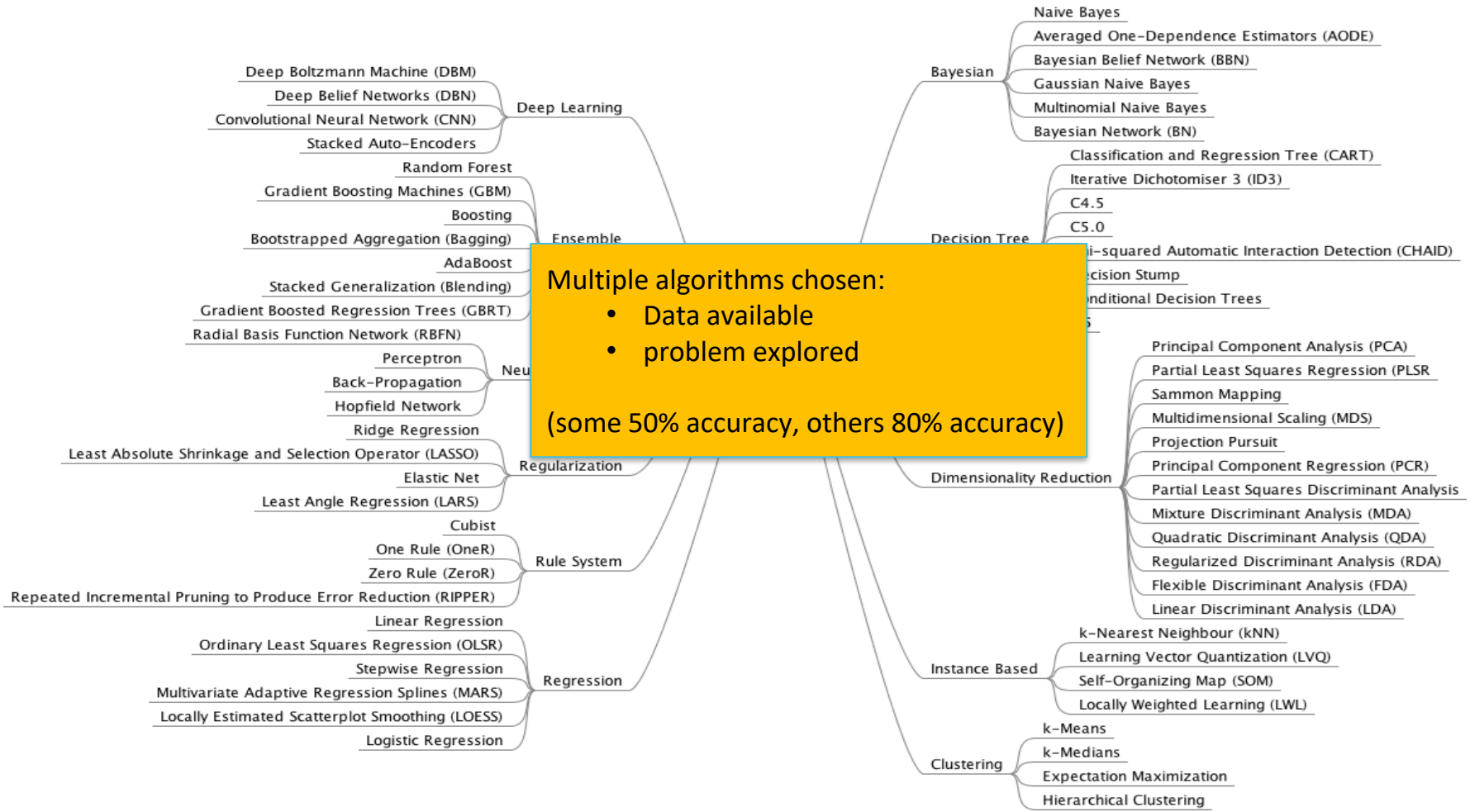






Example: ML in Action (classifying data)

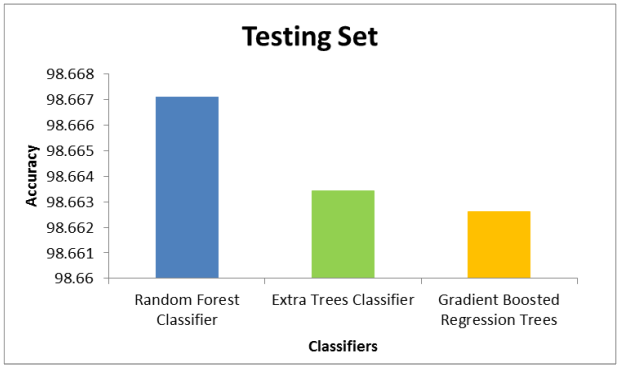


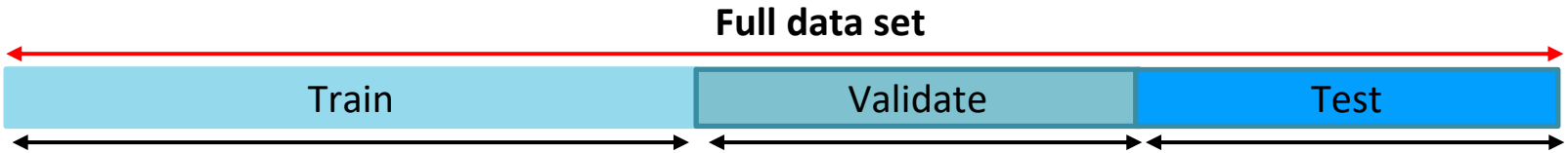

Multiple algorithms chosen:
• Data available
• problem explored
(some 50% accuracy, others 80% accuracy)
Deep Learning
Deep Boltzmann Machine (DBM)
Deep Belief Networks (DBN)
Convolutional Neural Network (CNN)
Stacked Auto-Encoders
Ensemble
Random Forest
Gradient Boosting Machines (GBM)
Boosting
Bootstrapped Aggregation (Bagging)
AdaBoost
Stacked Generalization (Blending)
Gradient Boosted Regression Trees (GBRT)
Neu
Radial Basis Function Network (RBFN)
Perceptron
Back-Propagation
Hopfield Network
Regularization
Ridge Regression
Least Absolute Shrinkage and Selection Operator (LASSO)
Elastic Net
Least Angle Regression (LARS)
Rule System
Cubist
One Rule (OneR)
Zero Rule (ZeroR)
Repeated Incremental Pruning to Produce Error Reduction (RIPPER)
Regression
Linear Regression
Ordinary Least Squares Regression (OLSR)
Stepwise Regression
Multivariate Adaptive Regression Splines (MARS)
Locally Estimated Scatterplot Smoothing (LOESS)
Logistic Regression
Bayesian
Naive Bayes
Averaged One-Dependence Estimators (AODE)
Bayesian Belief Network (BBN)
Gaussian Naive Bayes
Multinomial Naive Bayes
Bayesian Network (BN)
Decision Tree
Classification and Regression Tree (CART)
Iterative Dichotomiser 3 (ID3)
C4.5
C5.0
i-squared Automatic Interaction Detection (CHAID)
cision Stump
nditional Decision Trees
Dimensionality Reduction
Principal Component Analysis (PCA)
Partial Least Squares Regression (PLSR
Sammon Mapping
Multidimensional Scaling (MDS)
Projection Pursuit
Principal Component Regression (PCR)
Partial Least Squares Discriminant Analysis
Mixture Discriminant Analysis (MDA)
Quadratic Discriminant Analysis (QDA)
Regularized Discriminant Analysis (RDA)
Flexible Discriminant Analysis (FDA)
Linear Discriminant Analysis (LDA)
Instance Based
k-Nearest Neighbour (kNN)
Learning Vector Quantization (LVQ)
Self-Organizing Map (SOM)
Locally Weighted Learning (LWL)
Clustering
k-Means
k-Medians
Expectation Maximization
Hierarchical Clustering

Validating ML models: Measuring Accuracy (statistics)

Accuracy= (Number of correct predictions) / (Total number of predictions)

- Classifying flows over the network
  - Different models give different accuracies.
  - Different processing time (recurrent neural network (RNN) took 3 days to finish with 70% accuracy) (not shown)



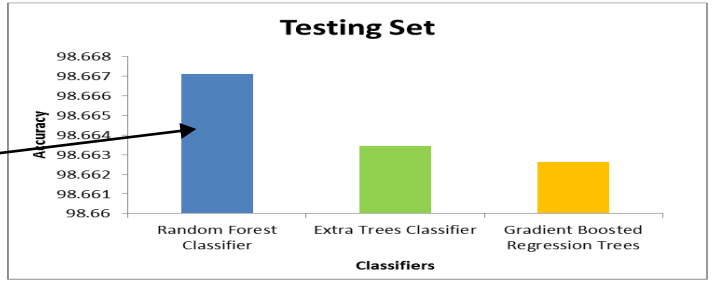


Model “trains”, e.g. 50% of data
Determines features, outputs

Check your model.
Model tuning,
estimate error rate

Unseen data set for testing:
Received accuracy

- Divide your data into 3 categories. Can also do 2 (train, test)
- Calculate accuracy of test data
- Justifies why you choose Random Forest



ML Libraries

- Scikit-learn good for learning, python library, free, lots of examples
- Use Jupyter Notebooks for investigating ML

Toolkit	Language	Application Use	Processing capability
Caffe	C++	Images and video	Distributed (HPC, GPU)
TensorFlow	Python	Images, regression, video, text, speech	Distributed (HPC, GPU)
Theano	Python	Images	Distributed (HPC, GPU)
Torch	Lua/Python	Images and speech	Distributed (HPC, GPU)

Explore Important Features in Netflow Data

Data Used

Supervised data:

- Jan 1st 2020: March 15th 2020 – PreCovid
- March 16th 2020: June 15th 2020 – PostCovid

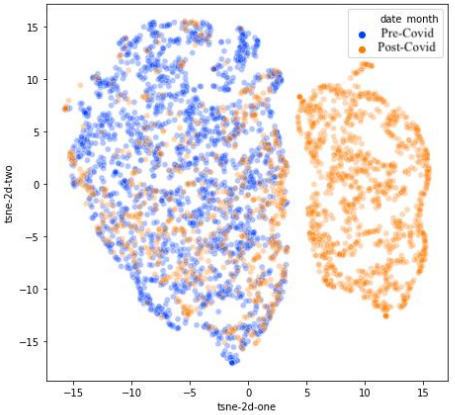
Type of feature (TCP(:6), UDP(:17), ICMP(:1))	Feature description
Bytes Inbound	Integer
Bytes Outbound	Integer
Packet Count Inbound	Integer
Packet Count Outbound	Integer
Unique Server IP Inbound recorded hourly	Integer
Unique Server IP Outbound recorded hourly	Integer
Unique Server Port Inbound recorded hourly	Integer
Unique Server Port Outbound recorded hourly	Integer

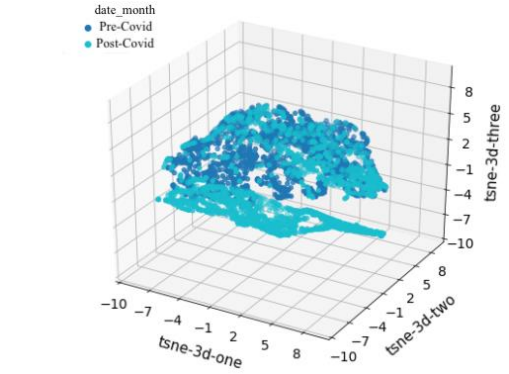
Example algorithms you can use:

- Decision Trees
- Random Forest
- PCA (Principal Component Analysis)
- tSNE, more....

Covid Impact on ESnet Netflow Data

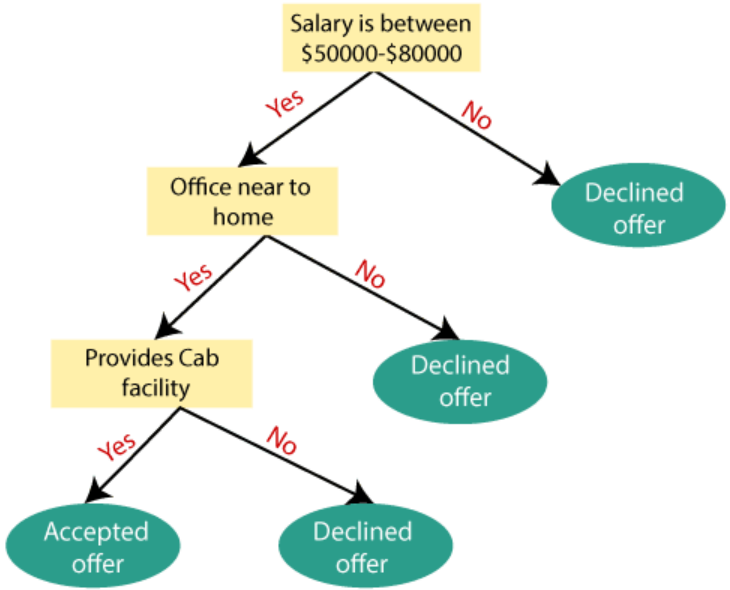
- Netflow records were summarized into hourly data – total number of flows, bytes, protocols etc.
- Multiple dimensions can be reduced into 2/3 dimensions using PCA or tSNE
- Good for quick visualization
- Need a few rounds to select optimal features for best clusters





Use Decision Trees (DT) to see what are the Dominant Features

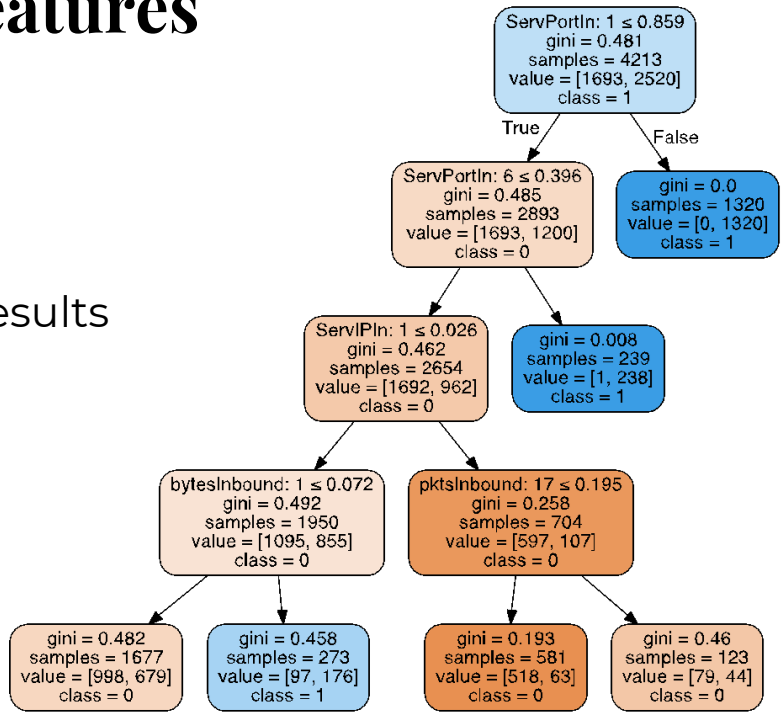
- Start with root node
- Based on rest of data grow the tree with 2 branches



- Information gain formula to help work out these leaf nodes. Works out a Gini Index as a probability one belongs to the class
- Prune the tree to optimal class representation (e.g. depth, leaf nodes)

Use Decision Trees (DT) to see what are the Dominant Features

- Rule-based classifier
- Simple and interpretable
- Fast
- Can lead to overfitting and bias in the results



Class 0 PreCovid

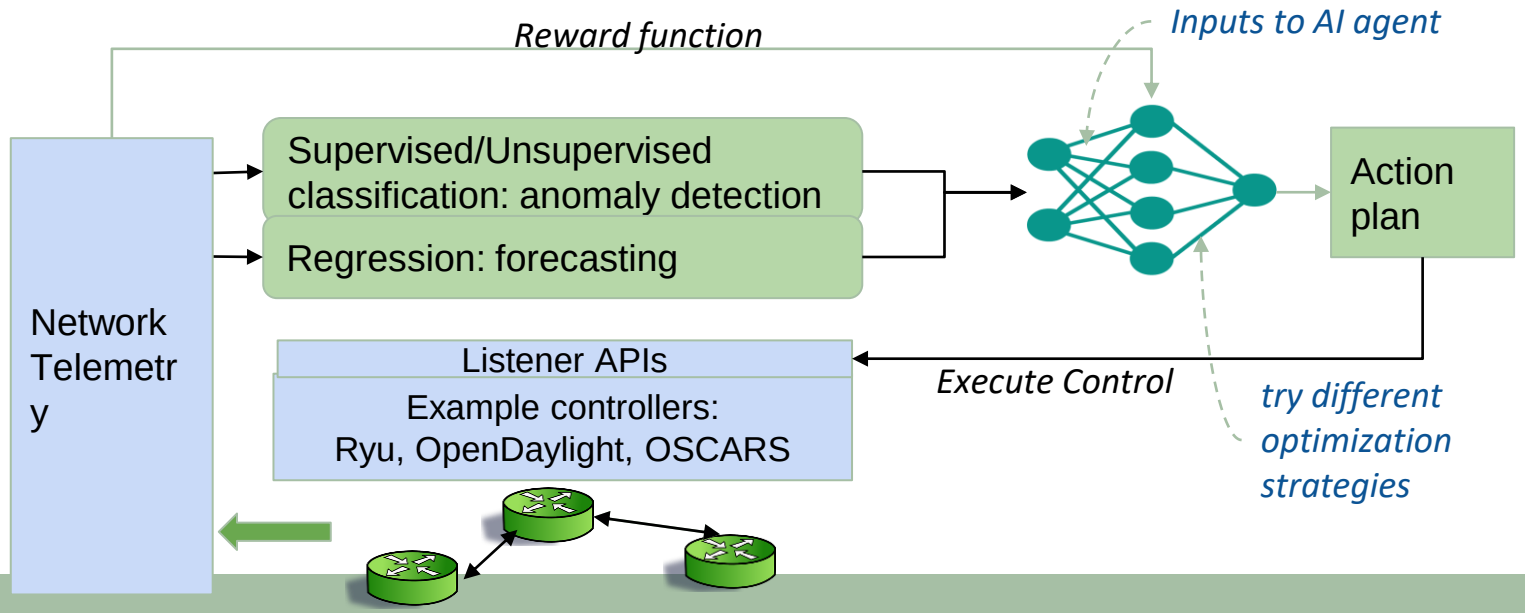
Class 1 PostCovid

e.g. Port numbers, unique IP addresses for logging in and Number of bytes moving in came up as dominant features

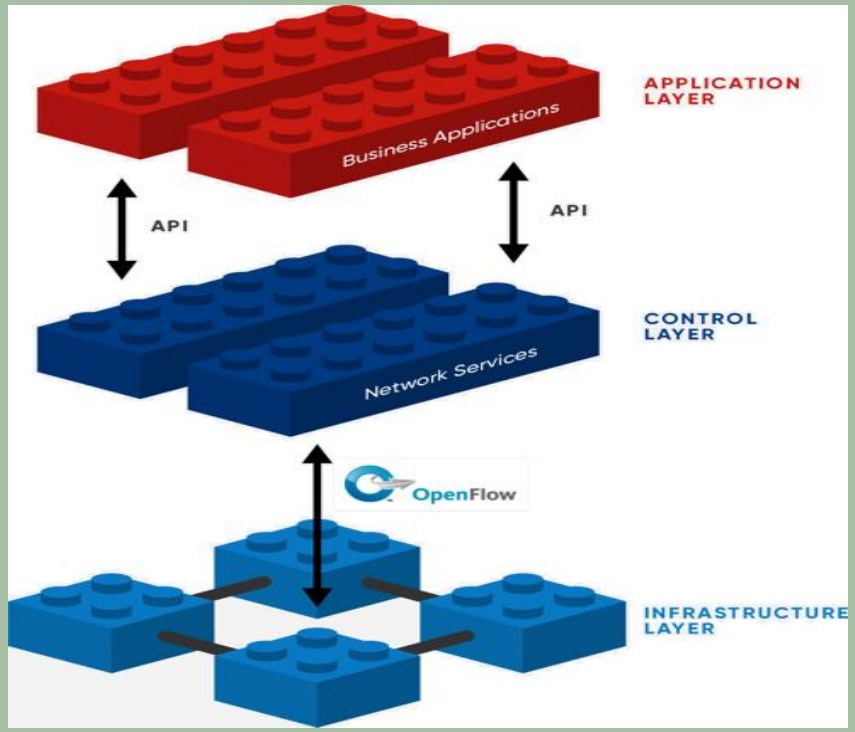
Conclusions Part 1

"Networks should learn to drive themselves"

simple actions such as improving availability,
attack resilience and dealing with scale



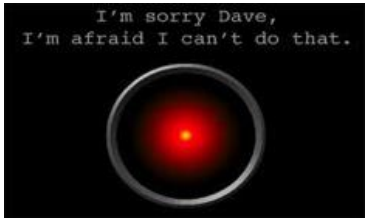
Why (and How) Networks Should Run Themselves, Feamster, Rexford



https://web.mit.edu/remy/
Ghosal et al. Model Predictive Congestion Control for TCP Endpoints
Lakhina et al. Diagnosing network-wide traffic anomalies

Software/Hardware Fusion Challenge

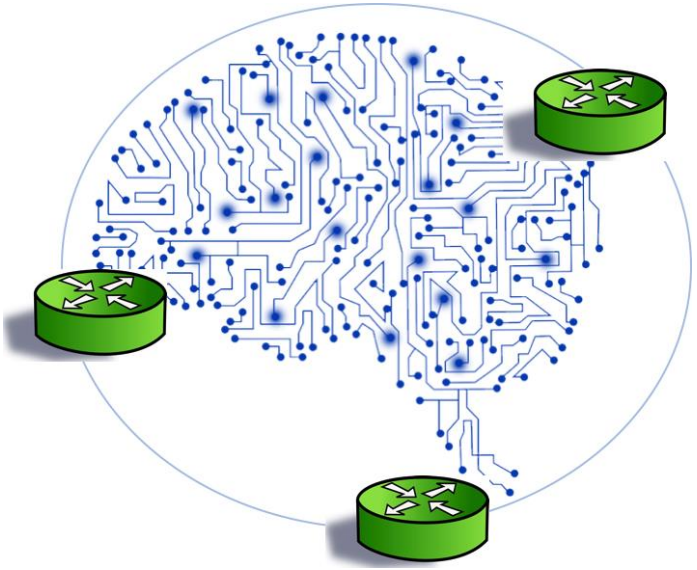
- Deep Learning trend
  - Image recognition e.g. Cats 2011
  - Self-playing game e.g. AlphaGo 2016
- Hardware acceleration
  - GPU advances
  - FPGAs
- Industry and Academic Efforts
  - Smart NICs e.g. Barefoot
  - AI @ Control Plane e.g. juniper, cisco)
  - AI enabled TCP
  - Traffic patterns



Next Topic:

Advanced Topics for using ML and P4 applications

Any questions?



Advanced Topics: Machine Learning and P4 Applications

Jun 2022
Introductory and Advanced Topics on P4 Programmable Data Plane Switches Workshop

Dr. Mariam Kiran, Prof Jorge Crichigno (UofSC)
<mkiran@es.net>
Scientific Networking Division
Lawrence Berkeley National Laboratory

BERKELEY LAB

U.S. DEPARTMENT OF ENERGY
Office of Science

Outline

P4 review

- Language
- Match-target cases
- How tos

P4 use cases discussions

Summary of ML applications with P4

Why P4?

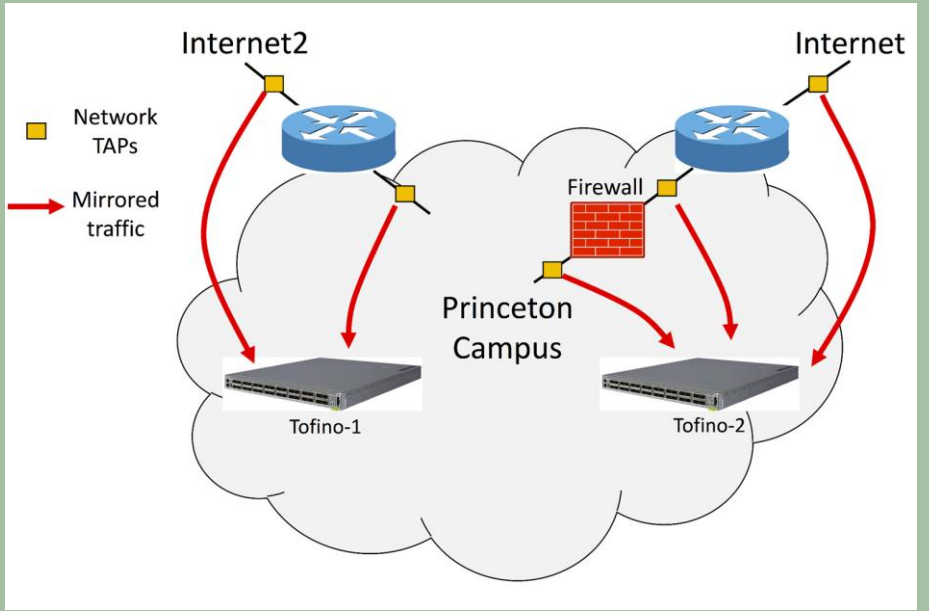
Flexibility

Cost

Faster

Efficient

towards bespoke networks?



https://p4campus.cs.princeton.edu/howto.html

A new flow arrives - what to do with it?

Example: In-band Telemetry from Q-Factor

Leveraging In-band Network Telemetry [2/3]

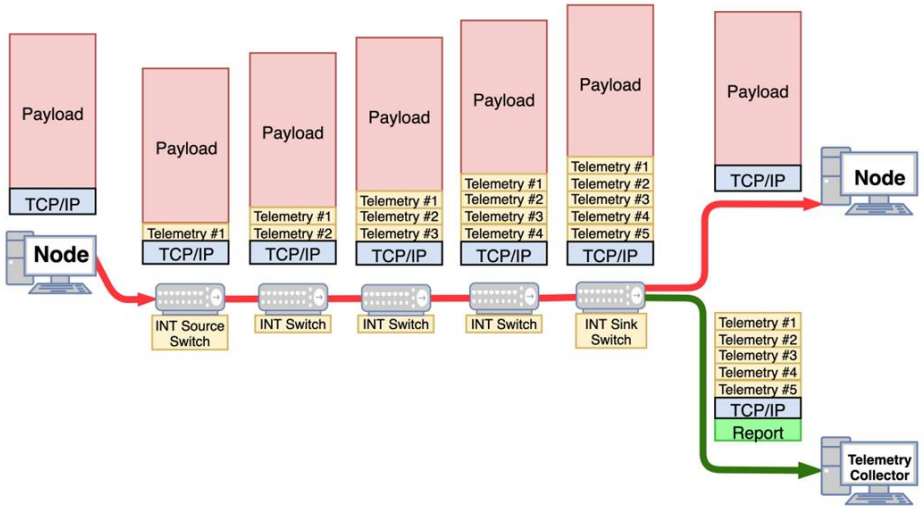
1 – User sends a TCP or UDP packet unaware of INT

2 – First switch (INT Source Switch) pushes an INT header + metadata

3 – Every INT switch pushes its metadata. Non-INT switches just ignore INT content

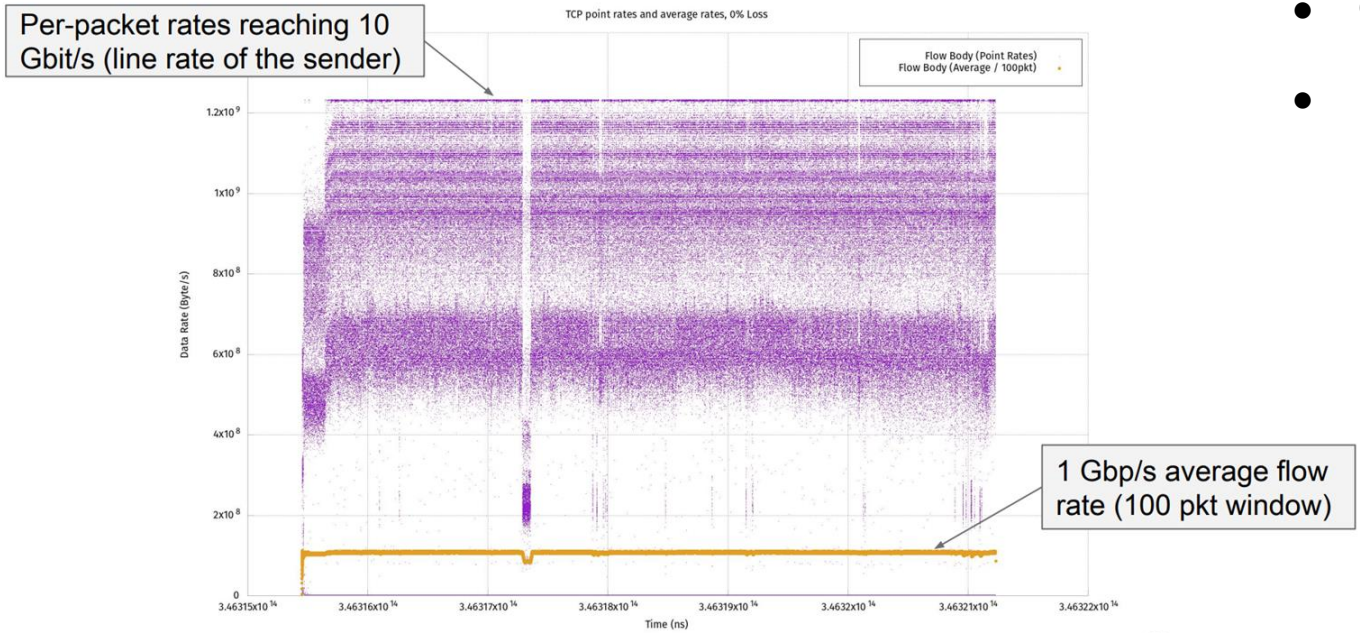
4 – Last switch (INT Sink Switch) extracts the telemetry and forwards original packet to destination

5 – Last switch (INT Sink Switch) forwards the 1:1 telemetry report to the Telemetry Collector



- Collects information as it moves through the network
- Collect in telemetry for analysis

Perform TCP analysis (Q-Factor)



- Challenges for data processing
- Built streaming pipelines using RabbitMQ

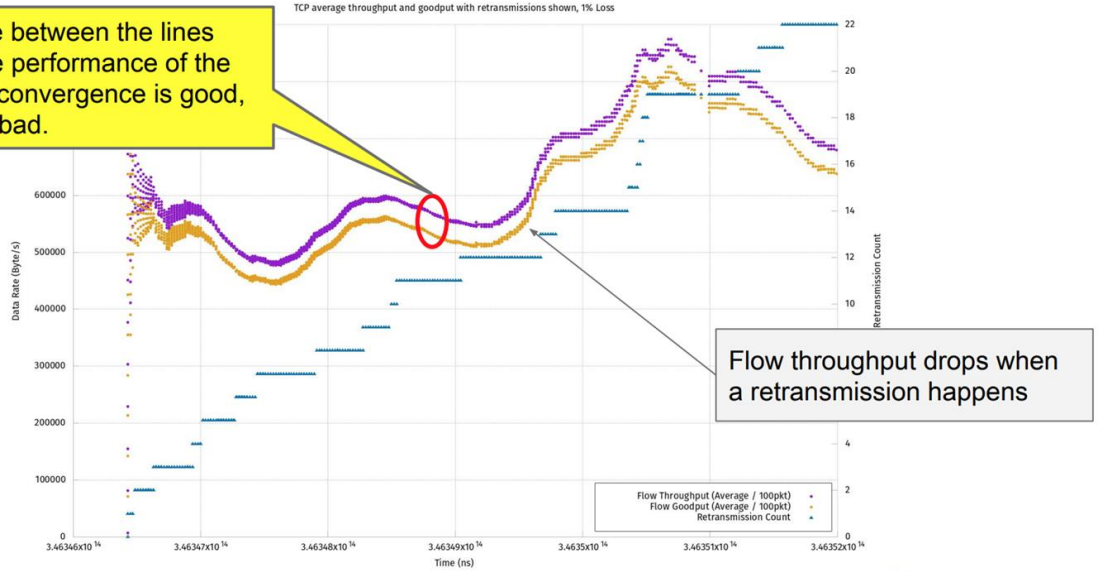
Note: Average rate is calculated using a time-weighted average of per-packet rates.

ESnet

TCP Analysis (2) - Q-Factor

1 Gbps iPerf flow - 1% packet drop - cont'd

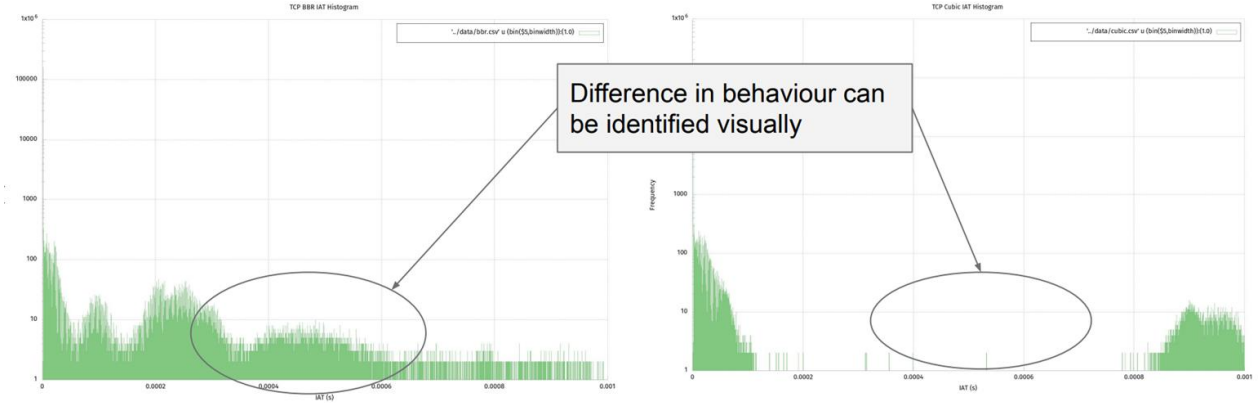
The difference between the lines represents the performance of the data transfer, convergence is good, divergence is bad.



ESnet

TCP Analysis (3) - Q- Factor

BBR vs Cubic - Inter-Arrival Time histogram



TCP BBR (delay-based)

TCP Cubic (loss-based)

BBR: inter-packet timing is more widespread than other congestion control algorithms.

ESnet

Working towards a Predictable Network

Common Approaches:

- Netflow
- SNMP
- Perfsonar
- TCP

Current Industry Applications

- Path tracing to check Correctness
  - e.g. in SDN reconfigure
- ECMP path checking problems

- Measurements
  - Build network predictions at finer granularity
  - Classification on data to find which TCP Congestion control is being used (i.e. LSTM models)
  - Finding anomalies

->Passive Analysis

What ideas would you try?