Self-Driving or *Autonomous* Networks

Dr. Mariam Kiran

Scientific Networking Division

Affiliations:
Computation Research Division, ESnet

*Lawrence Berkeley National Lab*
Self-Driving Technology (Real world and Fiction)

• Self driving cars (in Movies) can:
  – Drive themselves, through traffic, pick up and drop off
  – They can fly!

Other examples:
• Total recall
• Minority report
• And many more....
Movies/Comics good predictors of ‘Technology Hypes’!

• Science Fiction exploring it for ages
• Brought the main ideas around AI and human interactions
But In the Real World…. Companies gauging what to work on

- Gartner Hype Cycle
- Companies use this to chart the next ‘Big’ thing for commercial purpose
- Anything ‘hyped’ is always at the peak
- As technology matures, it becomes more reliable to work in
Five Levels of Autonomy (Cars)
ML and AI for Autonomy

• Object Detection
• Pattern Recognition
• Text mining
• Prediction systems
• Evidence-based systems
• Recommendation systems
• And more…..

• Artificial Intelligence (AI) vs Machine learning (ML) vs Deep Learning (DL)?
Difference between AI, ML and DL

- Turing’s paper “Can Machines Think!” – Turing Test: Exhibit human-like intelligence
  - Recently seen in movies

- 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
ML is a subset of AI

- Optimization techniques
- Expert systems
- Fuzzy systems
- Neural Networks
- Many more....
- Evolutionary algorithms (Genetic algorithms, evolutionary strategies, etc)
- Swarm intelligence (ant colony, particle swarm, more)
- Networks: graph algorithm (routing – shortest path)

Where ever learning involved (training): ML
Each algorithm is chosen depending on data being explored and problem being explored (some 50% accuracy, others 80% accuracy)
## Choosing Algorithms for Specific Problems

<table>
<thead>
<tr>
<th>Deep neural network</th>
<th>Input Data</th>
<th>Applied for</th>
<th>Variants</th>
</tr>
</thead>
</table>
| Feed forward neural network | Hierarchical data representations    | • Classification  
• Clustering  
• Anomaly finding  
• Feature extraction | • Deep belief networks  
(uses restricted boltzman machine for activation function)  
• Convolutional neural networks |
| Recurrent neural network  | Sequential data representation  
(i.e. time series data) | Sequential learning, especially useful when time relationship exists. | Long short term memory  
(LTSM) used for speech translation.                                     |

- There are many variants of DNNs. Papers and researchers in each specific DNN.
- DeepMind used Deep Q-learning for Attari and Go.
  - Action-pairs based on learned data.
Why Cars inspired us for Networks?

Similarity:
4 wheels, gears, motors, and more

Difference (some):
• Real-time monitoring dashboards
• ‘Softwarization’ of cars
• Automation
• Personalized service
Networks of the Future!

Similarity:
Switches, routers, links and devices

Difference (some):
• Real-time monitoring dashboards
• ‘Softwarization’ of networks
• Automation
• Personalized service

Monitoring logs and Machine Learning

Intent-based

Infrastructure-agnostic

Networking infrastructure

ESnet 6, 7, 8
Towards Autonomous Networks

With AI/ML/DL
ESnet Background

- R&E networks for science (CERN, LHC, and more)
- Provide reliable robust network connections to enable science workflows
- Investigate research and techniques to help build better networks
- Guarantees for our scientists for network needs (users)
Many Actors, softwares, data, etc ....

Slot-die printing of Organic photovoltaics

Borrowed from E Dart
ESnet Team continually engaged

- Science workflows (using tools like NSI, OSCARS)
  - Multi-domain provisioning (setting up link across many networks)
- Transfer tools and protocols (using Globus) (TCP research)
  - Ease of use, Reliable
- R&E Networks support big data oriented services (using ScienceDMZ)
  - Dedicated Bandwidth on demand, loss free
  - Isolation
  - Monitoring (perfSonar, traffic, cybersecurity)
  - Network virtualization
- Network research
  - Virtualization, SDN, switches, routers, etc

Designing for
- Specific science cases
- End-users
- Network engineers
Why we need Network Research?
A Day in the Life of a Packet

Problems of:
capacity, real-time response,
jeopardizes science reliability,
and more
ESnet Traffic Volume Growing Exponentially
Managing Multiple Sites together

- Different traffic requirements
  - Quality of service, bandwidth, speed, time-based deliveries, etc
- Reliability and heterogeneity
- Continuous upgrades to hardware and software
Networks and ML relationship (IETF forums)

- Predict traffic peaks
- Network security:
  - Find anomalies for security threats
- Path optimization
  - Link utilization
  - Divert traffic to other paths
- Predict link failures or packet loss
- Understand/ predict user behavior
- Find hardware/software bugs

- All are Core Network Research Problems!
Networks are Huge and Complex

- ESnet is a Wide Area Network (WAN) with multiple layers
- Current industries focus on specific case studies
- Over 2000 papers in the area
Behind the scenes: What does a Network look like?
Using ML for...

User traffic data
- User traffic (directed flows)

WAN Topology
- WAN Topology (traffic engineering)
  - (flow-level, traffic prediction, adaptation, path optimization, link failure)

Infrastructure traffic data
- Infrastructure-level modifications (Switches, deployment, etc)

- (Packet-level, queues, TCP, UDP)
What is Most Published?

- Most ML techniques used for classification (of traffic) and prediction (failures)
- Recent Google papers have been most influential:
  - B4, Jupiter, BwE, etc. (data center to user-based provisioning)
- Network tools enhanced by embedding informed decisions such as traffic awareness for:
  - Forming topologies, optimum path finding
  - Improve path utilizations depending on traffic
Why is ML research for Networks different?

- Complete Engineering problem (similar to car parts)
- Highly dynamic in nature
- Users are humans with many and diverse demands
- Multiple data sets and multiple devices to control
- ML for time-series data not Images
- React quickly to happening events (e.g. cybersecurity)
- Humans (engineers) have to be part of any ML solution
To Achieve Autonomy, building ML solutions

1. Anomalies in link performance: ARIMA
   - ANOMALY!

2. Classifying flows across DOE sites: Gaussian Mixture Models

3. Predicting traffic topologies across DOE sites: Markov Models
To Achieve Autonomy, building ML solutions (2)

4

Normal and abnormal transfers: PCA Feature extraction

Transfers with loss, packet duplication and reordering

Normal transfers

5

Predicting traffic per link/site: LSTM and 2-way encoders

Training input

Training output

Sliding Window
Building an Autonomous Network

Unrelated and diverse data sets across the WAN network

Unsupervised Feature Extraction and Deep learning

Feature extraction (object detection)
Classification
Clustering
Statistical Analysis

Regression
Prediction
Translation to code and take possible actions

Optimization and Automation of mundane tasks

DATA
Machine Learning
Translation to Networks

SNMP
Bro logs
Tstat
Netflow
Perfsonar
Tickets
Goal is to achieve **Autonomous Behavior**... not just ML in Networks

Intent-driven networks: INDIRA
Self-healing networks
Bringing it to Five levels of Autonomy for Networks

Every router, switches configured

Network recognizes needs and optimizes

Network senses something is wrong and corrects it

Intent-based Research

Self-driving Network
Intent-driven Networks: Setting the Stage

- Applications have complex workloads
- Network behavior tailored for my application ‘**intent**’
- Difficult to fulfill these diverse set of needs
  - Learning curve is huge and complex
  - Difficult to specify needs in ‘english’
  - Specify in high-level language, portable, multi-domain

---

I want to watch a movie tonight on Netflix

I want to see my real-time high-resolution big data visualization

I want to stream the big data directly into the cache of my super computer
Introducing iNDIRA... “Hello! I’m iNDIRA”

“I want to send data to my SuperComputer at NERSC by 5:00pm today”

“Ok I’ll reconfigure the network to make this possible!”

- Understand English (e.g. transfer, connect)
- Check conditions
- Ask for any further details
- Check conflicts and permissions

- Automate rendering into network commands like bandwidth, time schedule, topology
- Optimize the network
- Return success or failure to user
Setting Up Paths for Individual: **Intent**

- Traffic paths provisioned with basic QoS values, what if this is optimized for ‘end-users’
- Rather than running the following command: (setting up a link with QoS)

```
./onsa reserveprovision -g urn:uuid:6e1f288a-5a26-4ad8-a9bc-eb91785cee15
-d es.net:2013::bnl-mr2:xe-1_2_0:+#1000 -s es.net:2013::lbl-mr2:xe-9_3_0:+#1000
-b 5096
-a 2016-11-13T00:00:00 -e 2017-04-04T17:00:00 -u https://nsi-aggr-west.es.net:443/nsi-v2/ConnectionServiceProvider -p es.net:2013:nsa:nsi-aggr-west -r canada.eh:2016:nsa:requester -h 198.128.151.17 -o 8443 -l /etc/hostcert/muclient.crt -k /etc/hostcert/muclient.key -i /etc/ssl/certs/ -y -x -z -q;
```

- Networks can **understand** users: “Tell me what do you want!”
  - Example:

  **Scientist**> Can you set up a connection between Berkeley and Argonne.
  **Network**> Do you want guaranteed bandwidth?
  **Scientist**> Sure!
  **Network**> OK! Ill get this setup for you........................................ You’re all set!
Indira’s working: top level view

- Intent parser engine
- ML to optimize decisions
- Knowledge base
  - Project profiles
  - Semantics
  - Network services and conditions

- Backend details:
  - Multi-layer provisioning
  - Multi-tool connectivity
  - No platform dependence (multiple programming languages)

User interface

Intent-based Research

Virtual layer

Physical layer

Topology information

User profiles

Intent schema
Negotiate with Users

INDIRA> Hello and welcome! Tell me your project name?
Scientist> Bob
INDIRA> Is ‘Bob’ your project username? Y/N
Scientist> N
INDIRA> What is your project username
Scientist> Project1
INDIRA> Tell me what want the network to do? Hint “I want to [connect] [disconnect] services
Scientist> I want to transfer files from LBL to ANL
INDIRA> Do you have any time constraints with the transfer from LBL to ANL?
Scientist> No

INDIRA> Ok. Give me the link to the file to be moved.
Scientist> project_test@LBL/datasource1 to project2@ANL/dataset
INDIRA> Do you have any bandwidth requirements?
Scientist> Yes Maximum
INDIRA> Sorry, you have only 5Mb/s allowable to use. Do you want to use all of this?
Scientist>No
INDIRA> Thank you. Please contact the network administrator to allow you more bandwidth, as you don’t have permission to go above 5Mbps
Negotiate with Users

INDIRA> Hello and welcome! Tell me your project name?
Scientist> Bob
INDIRA> Is ‘Bob’ your project username? Y/N
Scientist> N
INDIRA> What is your project username
Scientist> Project1
INDIRA> Tell me what want the network to do? Hint “I want to [connect] [disconnect] services
Scientist> I want to transfer files from LBL to ANL
INDIRA> Do you have any time constraints with the transfer from LBL to ANL?
Scientist> No
INDIRA> Ok. Give me the link to the file to be moved.
Scientist> project_test@LBL/datasource1 to project2@ANL/dataset
INDIRA> Do you have any bandwidth requirements?
Scientist> No
INDIRA> Thankyou. I am configuring your transfer to start ‘now’ at 5GB/s...

........

Congratulations....All Done!
Self-driving Networks (infrastructure)

- ML algorithms explored across multiple layers
  - Behavior forecast and anomaly using DL
  - Simple classification for traffic patterns
- Recovery phase: Reactive or rule based systems (if.. then)
- Bring it all together to solve one problem
- Objective: Keeping network as ‘stable’ as possible
Open Research Challenges

• Maintaining Network Reliability
• Suitable machine learning algorithms
  – Real-time anomaly detection: Need quick response
  – Time based data
  – Different data collection issues: 1s versus 30s intervals
• Improve training time
• Engineering challenges:
  – Tools: What tools or devices we have control over to help automate recovery?
  – Scaling: Cost of processing data, quicken processing so that we can react quickly?
Summary

• Goal is to achieve ‘Autonomy levels’ in Network, eventually **Level 5**
• AI has shown promise: Used in a lot of applications in various fields
• Our efforts are focused on two main themes:
  – **Automation** and **Optimization**
• ESnet is at forefront of network (research, data, tools, expertise and complex apps) and Network/AI/ML research
• Leveraging on-site expertise and facilities NERSC, Lawrencium, more (also Google Cloud, Amazon EC2)
• Combining techniques (and algos) to advance research in explored:
  – New areas in Network Research!
Contact

• Thankyou!

• Lots of opportunities to engage:
  – Summer internships/students
  – Part-time/Full-time opportunities
  – Just come along and chat

• Feel free to reach out for more information/collaboration/ideas:
  
  • <Mkiran@es.net>