O projeto INSaNE: uma parceria entre Universidades, uma Startup e um ISP

Edmundo de Souza e Silva¹

Universidade Federal do Rio de Janeiro
¹Programa de Engenharia de Sistemas e Computação, COPPE

2019
Colaboradores

- **UFRJ:**
  - Rosa M.M. Leão, Daniel S. Menasche, Edmundo de Souza e Silva
  - 3 estudantes de doutorado, 2 estudantes de mestrado
  - Doutorado: Ananda Streit, Gabriel Mendonça, Gustavo Santos

- **UMass:** Don Towsley + PhD student Amir Reza Ramtin

- **Startup (incubada na COPPE/UFRJ):** Anlix

- **ISP:** Gigalink

E. de Souza e Silva
Parceria Universidade-Empresa, 2019
Measurement, modeling and analysis have been essential areas of research since the dawn of the Internet.

UCLA was the ARPANET Network Measurement Center.
Measurement is the first step that leads to control and eventually to improvement. If you can’t measure something, you can’t understand it. If you can’t understand it, you can’t control it. If you can’t control it, you can’t improve it.
Focus: lightweight measurements performed at home gateway routers.

... and data analysis.

Home routers are conduit to home devices → ideal place to implement measurement functionalities.

Project: collaboration among Brazil and the USA (UFRJ and UMass), Gigalink and Anlix incubated at UFRJ for data gathering and analysis.
Objective

Focus: lightweight measurements performed at home gateway routers.

... and data analysis.

Home routers are conduit to home devices → ideal place to implement measurement functionalities.

Project: collaboration among Brazil and the USA (UFRJ and UMass), Gigalink and Anlix incubated at UFRJ for data gathering and analysis.
Objective

- Focus: lightweight measurements performed at home gateway routers.
- ... and data analysis.
- Home routers are conduit to home devices → ideal place to implement measurement functionalities.
- Project: collaboration among Brazil and the USA (UFRJ and UMass), Gigalink and Anlix incubated at UFRJ for data gathering and analysis.
Focus: lightweight measurements performed at home gateway routers.

... and data analysis.

Home routers are conduit to home devices → ideal place to implement measurement functionalities.

Project: collaboration among Brazil and the USA (UFRJ and UMass), Gigalink and Anlix incubated at UFRJ for data gathering and analysis.
Three issues:

**Issue 1:** Automatically detect problems in the network.

**Issue 2:** Assess quality of experience (QoE) of residential users.

**Issue 3:** Analyze traffic: try to find traffic profiles.

**Issue 4:** investigate DDoS detection at the network edge.
Three issues:

**Issue 1:** Automatically detect problems in the network.

**Issue 2:** Assess quality of experience (QoE) of residential users.

**Issue 3:** Analyze traffic: try to find traffic profiles.

**Issue 4:** Investigate DDoS detection at the network edge.
Objective

Three issues:

**Issue 1:** Automatically detect problems in the network.

**Issue 2:** Assess quality of experience (QoE) of residential users.

**Issue 3:** Analyze traffic: try to find traffic profiles.

**Issue 4:** Investigate DDoS detection at the network edge.
Three issues:

**Issue 1:** Automatically detect problems in the network.

**Issue 2:** Assess quality of experience (QoE) of residential users.

**Issue 3:** Analyze traffic: try to find traffic profiles.

**Issue 4:** Investigate DDoS detection at the network edge.
Measurement Infrastructure
Partnership University + ISP + startup at COPPE/UFRJ

- Data collection campaign at the network edge
- Almost 4,000 homes collecting data (and growing)
Measurement Infrastructure

- Partnership University + ISP + startup at COPPE/UFRJ
- Data collection campaign at the network edge
- Almost 4,000 homes collecting data (and growing)
Measurement Infrastructure

- Partnership University + ISP + startup at COPPE/UFRJ
- Data collection campaign at the network edge
- Almost 4,000 homes collecting data (and growing)
Measurement Infrastructure

Data

- Traffic (byte counts and packet count): download/upload
- loss rate, latency
- processor load, etc.
- WiFi: transmission rate, SNR, etc.
- measurement intervals: 1 minute: Large number of time series
**Measurement Infrastructure**

**Data**

- **Traffic (byte counts and packet count): download/upload**
- **loss rate, latency**
- **processor load, etc.**
- **WiFi: transmission rate, SNR, etc.**
- **measurement intervals: 1 minute: Large number of time series**

![Diagram of network setup with a home router connecting a phone, TV, laptop, and server through the internet.]
Measurement Infrastructure

Data

- Traffic (byte counts and packet count): download/upload
- loss rate, latency
- processor load, etc.
- WiFi: transmission rate, SNR, etc.
- measurement intervals: 1 minute: Large number of time series
Measurement Infrastructure

Data

- Traffic (byte counts and packet count): download/upload
- loss rate, latency
- processor load, etc.
- WiFi: transmission rate, SNR, etc.
- measurement intervals: 1 minute: Large number of time series
Measurement Infrastructure

Data

- Traffic (byte counts and packet count): download/upload
- loss rate, latency
- processor load, etc.
- WiFi: transmission rate, SNR, etc.
- measurement intervals: 1 minute: Large number of time series
Example: data collection infrastructure
Goals

- Perform extensive **data analysis**: Machine Learning techniques
- Apply unsupervised learning techniques of spatial/temporal patterns to identify anomalous behavior.
- Identify changes (traffic, etc.) and automatically infer possible causes of these changes.
- New perspectives to understanding behavior of domestic networks
Goals

- Perform extensive **data analysis**: Machine Learning techniques
- Apply unsupervised learning techniques of spatial/temporal patterns to identify anomalous behavior.
- Identify changes (traffic, etc.) and automatically infer possible causes of these changes.
- New perspectives to understanding behavior of domestic networks
Goals

- Perform extensive data analysis: Machine Learning techniques
- Apply unsupervised learning techniques of spatial/temporal patterns to identify anomalous behavior.
- Identify changes (traffic, etc.) and automatically infer possible causes of these changes.
- New perspectives to understanding behavior of domestic networks
Goals

- Perform extensive **data analysis**: Machine Learning techniques
- Apply unsupervised learning techniques of spatial/temporal patterns to identify anomalous behavior.
- Identify changes (traffic, etc.) and automatically infer possible causes of these changes.
- New perspectives to understanding behavior of domestic networks
A lightweight approach for DDoS detection at home gateways
Objective

- DDoS attacks are prevalent! more than 1/3 of class C active networks were targeted by this kind of attacks (2015 and 2017)
  
  “… despite the law enforcement efforts, DDoS attacks remain a real threat to business”, DDoS report May 21, 2019.
- How to detect problems and react fast?
DDoS detection

- **Simplicity**: Methodology – simple enough to scale;
- **Locality**: Detection of problems close to the source;
- **Efficiency**: Fast attack identification;
- **Minimalism**: Collect a minimum amount of information at the home-routers. (Avoid interfering with performance and invading users’ privacy.)
DDoS detection

- **Simplicity**: Methodology – simple enough to scale;
- **Locality**: Detection of problems close to the source;
- **Efficiency**: fast attack identification;
- **Minimalism**: collect a minimum amount of information at the home-routers. (avoid interfering with performance and invading users’ privacy.)
**Simplicity**: Methodology – simple enough to scale;

**Locality**: Detection of problems close to the source;

**Efficiency**: fast attack identification;

**Minimalism**: collect a minimum amount of information at the home-routers. (avoid interfering with performance and invading users’ privacy.)
DDoS detection

- **Simplicity**: Methodology – simple enough to scale;
- **Locality**: Detection of problems close to the source;
- **Efficiency**: fast attack identification;
- **Minimalism**: collect a minimum amount of information at the home-routers.
  (avoid interfering with performance and invading users’ privacy.)
**Lightweight Approach For DDoS detection**

- **Question:** Is it feasible to detect DDoS attacks, in an extremely lightweight fashion, *without relying on information from packet headers*?

- **Methodology**
  - Machine learning techniques!
Lightweight Approach For DDoS detection

**Question:**
Is it feasible to detect DDoS attacks, in an extremely lightweight fashion, *without relying on information from packet headers*?

**Methodology**
- Machine learning techniques!
A few Results
Example: Classification Results

<table>
<thead>
<tr>
<th>Botnet</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mirai/BASHLITE</td>
<td>0.999688</td>
<td>0.997151</td>
<td>0.992361</td>
<td>0.994751</td>
</tr>
</tbody>
</table>

- Different types of attacks (e.g. TCP ACK flood, etc.). Real attack code.
- Considered different ML models
- Detecting DDoS attacks requires identification of subtle patterns in the baseline home-user traffic.
- Did not use packet headers
A change point detection problem

Inferring network problems and QoE from the edge
Inferring network problems from the edge

Objective:
- automatic problem identification: Congestion, equipment problems, attacks
- Temporal pattern identification from data collected

Issues:
- Is the user QoE affected?
- Can we correlate QoS metrics with technical call to ISP call center?
Inferring network problems from the edge

**Objective:**
- automatic problem identification:
  - Congestion, equipment problems, attacks
- Temporal pattern identification from data collected

**Issues:**
- Is the user QoE affected?
- Can we correlate QoS metrics with technical call to ISP call center?
Inferring network problems from the edge

Methodology

- **Machine learning** approaches for detecting and locating network problems.
- Construct models to detect statistical changes: *Change Point Detection* problem.
- Spatial-time correlations using ISP network topology.
- Locate root causes
- QoE
Inferring network problems from the edge
Dataset and model

- 2485 clients between 08/08/2018 e 23/09/2018
- Packet loss model.
- 51041 time series for training.
  25700 time series for testing.
Inferring network problems from the edge
Dataset and model

- 2485 clients between 08/08/2018 e 23/09/2018
- Packet loss model.
  - 51041 time series for training.
  - 25700 time series for testing.
Inferring network problems from the edge
Dataset and model

- 2485 clients between 08/08/2018 and 23/09/2018
- Packet loss model.
- 51041 time series for training.
  25700 time series for testing.
Inferring network problems from the edge
Change point detection problem
Inferring network problems from the edge
Inferring network problems from the edge
Inferring network problems from the edge
Inferring network problems from the edge
Additional Results

- Considering phone calls to the Call Center (technical complaints), associated with nodes connected to home-router performing measurements: The method identified 95% of phone calls.
Considering phone calls to the Call Center (technical complaints), associated with any nodes in topology: The method identified 89% of phone calls.
Learning Traffic Profiles from the Edge
Learning Traffic Profiles from the Edge

Objective:
- Understand characteristics of traffic generated by home users
- How to extract meaningful features from dataset, preserving users’ privacy (e.g., assuming that traffic is fully encrypted)

Methodology: based on Machine learning techniques
Learning Traffic Profiles from the Edge

**Objective:**
- Understand characteristics of traffic generated by home users
- How to extract meaningful features from dataset, preserving users’ privacy (e.g., assuming that traffic is fully encrypted)

**Methodology:** based on Machine learning techniques
Learning Traffic Profiles from the Edge

- Median download and upload traffic per minute for all user-day of each cluster.
Learning Traffic Profiles from the Edge

- User profile **for a month**.
  
  (A, B, C, D, E are clusters - *user-day* pattern).
Podemos aprender muito com medições nos roteadores residenciais e com *muito pouca informação* e técnicas de inteligência artificial.

- Detecção de problemas na rede antes do usuário
- Inferir QoE dos usuários
- Detectar padrões de tráfego do usuários
- Detectar ataques DDoS
- E mais …

Parceria Universidade/ISP/Startup:
- Produtos na fronteira do conhecimento
- Produtos voltados para necessidades dos ISPs
- Produtos bem testados no laboratório e na empresa parceira
Podemos aprender muito com medições nos roteadores residenciais e com *muito pouca informação* e técnicas de inteligência artificial.

- Detecção de problemas na rede antes do usuário
- Inferir QoE dos usuários
- Detectar padrões de tráfego do usuários
- Detectar ataques DDoS
- E mais . . .

**Parceria Universidade/ISP/Startup:**

- Produtos na fronteira do conhecimento
- Produtos voltados para necessidades dos ISPs
- Produtos bem testados no laboratório e na empresa parceira
Podemos aprender muito com medições nos roteadores residenciais e com *muito pouca informação* e técnicas de inteligência artificial.

- Detecção de problemas na rede antes do usuário
- Inferir QoE dos usuários
- Detectar padrões de tráfego do usuários
- Detectar ataques DDoS
- E mais . . .

Parceria Universidade/ISP/Startup:

- Produtos na fronteira do conhecimento
- Produtos voltados para necessidades dos ISPs
- Produtos bem testados no laboratório e na empresa parceira
Obrigado

OBRIGADO

PERGUNTAS?

www.abc.org.br/~edmundo

www.land.ufrj.br/~edmundo

anlix.io

www.land.ufrj.br/~classes/

machine-learning-2019/