Hands Off the Wheel in Autonomous Vehicles? 
A Systems Perspective on over a Million Miles of Field Data

Subho S. Banerjee, Saurabh Jha, James Cyriac, Zbigniew T. Kalbarczyk and Ravishankar K. Iyer

Computer Science, Electrical and Computer Engineering
Resilience of Autonomous Vehicles

- AVs advertised as transformative – improve congestion, safety, productivity, and comfort.

- Recent media attention on Tesla/Waymo/Uber AVs.

- Research Gap: Resilience of AV Technology
  - Causes – Dynamics – Impacts of failure
Overview

Data driven analysis of failures in the field during testing of AVs

California Department of Motor Vehicles AV Testing Reports (2014 – 2016)
1,116,605 miles – 144 AVs – 12 Vendors
5328 Disengagements – 42 Accidents

Failure Modes

1 Disengagements

Disengagement: A transfer of control from the autonomous system to the human driver in the case of a failure.

Accident: An collision with other vehicles, pedestrians, or property.

Quantified in terms of disengagements per mile (DPM) and accident per mile (APM).

2 Accidents

Human Initiated

AV Initiated
Key Findings

• AVs are up to 4000× more likely to have an accident than human drivers.

• DPM, APM strongly negatively correlated with miles driven.

• ML components of AVs responsible for 65% of failure reports.

• Reliability per mission: AVs are up to 100× worse than airplanes.
LogDriver: An End-to-End Workflow for AV Log Data Analysis

Stage I: Data Collection
- Vendors are required to collect data as per CA laws.
- CA DMV curates databases of vendor reports.
- No standardized reporting formats.
- Reports stored as scanned documents.
- Vendor specific parsing & filtering.
- Standardizing data formats across vendors.
- Data from 2016 and 2017 reports
- 5,328 disengagements
- 42 accidents
- 1,116,605 autonomous miles driven
- Each accident is reported separately

Stage II: Parsing and Filtering
- Parse + Filter
- Normalize
- OCR

Stage III: Natural Language Processing
- Parse natural language text relating to failure causes.
- Localize failures in abstract system model.
- NLP

Stage IV: Statistical Analysis of Failures
- Analyze failure data to quantify
  - Causes
  - Dynamics
  - Impacts

CONSOLIDATED
FAILURE
DATA
LogDriver: Nissan Case Study

1 Nissan Disengagement Reports from the CA DMV

2 Individual Report

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Location</th>
<th>Event Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/25/2016</td>
<td>11:20am</td>
<td>Leaf #1 (Alfa)</td>
<td>The AV didn't see the lead vehicle, driver safely disengaged and resumed manual control.</td>
</tr>
</tbody>
</table>

3 OCR + Parsing + Cleaning

The AV didn’t see the lead vehicle...

4 STPA[1] based ontology model

Categories: Recognition

## Data Driven Insights

<table>
<thead>
<tr>
<th>Topic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maturity of AV Technology</td>
<td></td>
</tr>
<tr>
<td>Causes of Failures</td>
<td></td>
</tr>
<tr>
<td>Improvement in AV Technology over Time</td>
<td></td>
</tr>
<tr>
<td>Hands Off the Wheel?</td>
<td></td>
</tr>
<tr>
<td>Safety: AVs vs Humans</td>
<td></td>
</tr>
</tbody>
</table>
Maturity of AV Technology

- DPM related to cumulative miles driven.
- Maturity: Still in “burn-in” phase.
Causes of Failures

- **ML/Design issues** responsible for 65% of failures.
- 48% of disengagements are **human initiated**.
Are AVs improving over time?

- Strong negative correlation of DPM with miles driven.
- Some manufacturers show increasing DPM trends
Hands off the wheel?

- Accident Avoidance Times less than non-AVs: 0.82 s (for AVs) vs 1.09 s (for non-AVs)
- 69% of reports accidents are “Latency Accidents”
Comparison to human drivers

- Non-AVs are $15 - 4000 \times$ less likely to have an accident.
- All accidents happen at intersection of urban streets.
- All accidents at low speeds: Human drivers cannot predict behavior.

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Median DPM (mile$^{-1}$)</th>
<th>Median APM (mile$^{-1}$)</th>
<th>Rel. to HAPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mercedes-Benz</td>
<td>0.565</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Volkswagen</td>
<td>0.0181</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Waymo</td>
<td>0.000745</td>
<td>$4.140 \times 10^{-5}$</td>
<td>20.7×</td>
</tr>
<tr>
<td>Delphi</td>
<td>0.0263</td>
<td>$4.599 \times 10^{-5}$</td>
<td>22.99×</td>
</tr>
<tr>
<td>Nissan</td>
<td>0.0413</td>
<td>$3.057 \times 10^{-4}$</td>
<td>15.285×</td>
</tr>
<tr>
<td>Bosch</td>
<td>0.811</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>GM Cruise</td>
<td>0.177</td>
<td>$8.843 \times 10^{-3}$</td>
<td>4421.5×</td>
</tr>
<tr>
<td>Tesla</td>
<td>0.250</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

HAPM – Human APM.
Human APM = $2 \times 10^{-6}$ mile$^{-1}$ [37], [38].

Column 4 = AV APM/Human APM.
Trends in 2017 Reports

• Vendors have moved away from testing vehicles in California.
  • BMW, Ford, Tesla, Honda, Volkswagen

• Decreasing DPM trend?
  • Not anymore...

• Serious issue:
  • Ridesharing as primary application.
  • Thousands vehicles.
  • $4.14 \times 10^{-5}$ DPM corresponds to multiple failures daily.
Looking Forward

• Functionality first => Resilience second

• AVs are here to stay
  • ML Perception/Decision Control is key culprit
  • Traditional reliability bugs (bit flips) seem less important
  • Foundation of new research thrust

• Need for new reliability metrics
Questions?