IEEE Emerging Technology Reliability Roundtable 2024

May 21-22, 2024

Lisbon, Portugal

https://cqr.committees.comsoc.org/etr-rt-2024/

Participants' Input to the Roundtable Summary of Findings

Introduction

The scope of the Emerging Technologies Reliability Roundtable (ETR-RT) is to:

- Discuss and identify the RAS (Reliability, Availability and Serviceability[1]) challenges, requirements, and methodologies in the emerging technology areas such as Cloud Computing, Wireless/Mobility (with focus on 5G technologies), NFV (Network Functions Virtualization), SDN (Software Defined Networking), or similar large-scale distributed and virtualization systems.
- Discuss the RAS requirements and technologies for mission-critical industries (e.g., airborne systems, railway communication systems, industrial automation, the banking and financial communication systems, etc.), health care systems, and autonomous vehicles with the goal to promote the interindustry sharing of related ideas and experiences.
- Investigate the use of ***large language models (LLMs)*** to find new/increase coverage of failure models – including security defects, operational errors and other failures.
- Identify ***Serviceability*** and ***Reliability-impacted Security aspects*** under ***Resilience***, which is considered an "umbrella" topic.

[1] Serviceability: Focus on the technology, solutions and platforms which can improve Operation and Maintenance (O&M) ability, including Deployment, Monitoring, Upgrade/Patching, Adjustment, Troubleshooting, Fault Recovery, Data Configuration, etc. The target of Serviceability is to make the O&M work more Easily, Remotely, Automatically and Efficiently to achieve Low OPEX and High Customer Satisfaction.

- Discuss the enormous impact of Artificial Intelligence (AI), with an emphasis on ***reliability/trustworthiness of AI solutions (Generative AI, Responsible AI, Sustainable AI[2], and AI for Sustainability [3]).***
- Identify potential directions for resolving identified issues and propose possible solutions.

In 2024 the focus was on:

- **IoT Reliability**
  - Reliability and serviceability for IoT applications, and massively scaled infrastructure
  - Reliability and serviceability for Digital Energy
  - Autonomous Vehicles Safety
- **AI Reliability and Responsibility**
  - Reliability and serviceability of massively scaled infrastructure for AI Foundation Models
  - Responsible and trustworthy AI for mission critical applications
  - Emerging regulations
- **Communications Heterogeneity Reliability**
  - Reliability and serviceability of Heterogeneous networks (e.g., 5G/WiFi6) which support mission critical applications including, Emergency Responders, Telemedicine, AR/VR/XR, Manufacturing/Industry 4.0 among others.
  - Identify challenges (e.g., Security and QoS) that impact resiliency of Heterogeneous networks and discuss remediation approaches.

Speakers/Participants and Talk Titles

This Roundtable was held on May 21-22, 2024 in Lisbon, Portugal. It was well balanced with speakers/participants from Telecom Service Providers (AT&T, Telefónica), Telecom Suppliers (Huawei, NOS Comunicações S.A.), IT companies (Google Research, Palindrome Technologies, Brighten AI, RoadToAmherst, LocalAI, INESC INOV-Lab), and universities (Czech Tech

[2] Sustainable AI: Refers to creating and using trustworthy AI systems that deliver long-term benefit to business, people, and the planet. It is most known as “Green AI” or “Eco-friendly AI”, what refers to the practice of developing and deploying AI technologies in a manner that ***minimizes their environmental impact*** and maximizes their long-term sustainability. It encompasses various strategies and principles aimed at reducing the carbon and ecological footprint of AI systems throughout their lifecycle, from design and manufacturing to operation and disposal.#

[3] AI for Sustainability: Refers to the use of artificial intelligence (AI) technologies to address environmental, ecological, and social challenges and promote sustainable development. It involves applying AI techniques to collect, analyze, and interpret data related to issues such as climate change, resource management, conservation, and other sustainability concerns.

Thus: Sustainable AI aims to make AI technologies more environmentally and socially friendly. In contrast, AI for Sustainability aims to leverage AI to directly advance environmental and social goals.

University in Prague, University. of Lisbon, University of Delaware, University of Illinois at Urbana-Champaign – UIUC, and University of Oulu).

The speakers addressed various issues and challenges of the 5G deployment and 6G research, the telecom regulatory environment, and the AI landscape providing further insight and impact on reliability, serviceability, and security. The talks provided "food for thought" questions in the common discussion held after each presentation and summarized in the closing discussion for highlighting major issues to address. Following are the titles of the presentations:

[#01] Spilios Makris – Introduction to ETR-RT2024

[#02] David Lu – ETR-RT2024 Opening Remarks

[#03] Lynette Webb – Update on AI Regulation

[#04] Kathy Meier-Hellstern – Scaling Responsible AI

[#05] John Burkey – Unleashing Potential Scaling AI for Impact (Talk 1)

[#06] Kostas Vlahodimitropoulos – AI for Energy Sustainability (Presentation)

[#06A] Kostas Vlahodimitropoulos – AI for Sustainable Energy (Paper)

[#07] Antonio Grilo – AI for Communications Networks

[#08] Pavel Kordik – Latest AI Developments Explained

[#09] David Guillen AI – Empowered Cloud Continuum for Mobile Networks

[#10] David Lu – Disruptive Impact 6G Future (Talk 1)

[#11] Matti Latva-aho – 6G Key Enabler for Metaverse

[#12] Lily Prasad – Data Driven Planning for Telecom under Climate Change

[#13] Luis Santo Green – Networks Minimizing Energy Consumption

[#14] Chengqiang Huang – Design for Dependability and its Challenges

[#15] David Lu – Risk and Opportunity of Large-Scale Network Outage (Talk 2)

[#16] Jun Xu – Understanding Battery Safety and Durability Issues

[#17] Tianyin Xu – Kubernetes Reliability and Reliable Cloud Infrastructures

[#18] Peter Thermos – Security and Reliability in Heterogeneous Networks

[#19] John Burkey – Harmonizing Realms (Talk 2)

[#20] David Lu – Hosting ETR-RT2025 in Prague Czech Republic**.**

After the Roundtable, participants were offered the opportunity to provide some further analysis, which is reported hereafter. The Roundtable speakers' and participants' inputs and additional comments well reflect the consensus that was discussed during the Roundtable, and they provided an interesting "food for thought" for future work in ETR-RT. They are listed below and show some similar concerns. Specifically, it was pointed out that:

The world is moving rapidly into an AI-driven frenzy and the speed of innovation accelerates in many areas such as Semiconductors, AI-based processors (GPU), Sensors, Quantum Computing, Low Earth Orbit Satellites, Low Air Economy (self-drive flying cars and drones), and regenerative medicine. At the same time, the geopolitical conflicts, and economic models (along with the technology advancement) drive further reliability and security challenges in an unprecedented pace. CQR as a premier engineering organization focused on reliability and its Emerging Technology Reliability Roundtable (ETR-RT) must stay ahead of the industry and promote new standards and models to address such challenges and ensure technology change brings positive impact to human societies.

Following are a few areas:

1. The current social and human interactions are changing dramatically compared to 5 or 10 years ago. What reliability and security concerns do we need to think about and address?
2. As our technology ecosystem is getting more complicated, the economic model drives profitability with never-ending thirst, how do we ensure the complexity of supply chain and end-to-end (ETE) reliability and security being addressed adequately? For example, Boeing's accelerated 737-Max development diminished the thorough testing required which resulted in disastrous accidents and impacted brand credibility, long-term profitability and sustainable growth! Of a greater concern is the manifestation of this issue across the supply chain where business decisions driven by monetary gain overrule sound engineering practices!
3. How do we balance divide-and-conquer vs. ETE ecosystem integration?
4. How could a user-centric view help improve industry reliability and security?
5. How do we best address cross-technology integration, such as terrestrial/satellite communications, as well as communication and sensor networks (hardware, and software)?
6. How will we, as the scientific and engineering community, ensure data integrity and accuracy, which is the critical element of delivering AI results?
7. How can we leverage traditional encryption technologies vs. emerging techniques (i.e., Blockchain, quantum computing), and/or the combination of all, to achieve zero-tolerance security?
8. Can we leverage generative AI to auto-generate and auto-execute testing cases to ensure higher product quality, reliability, and security?
9. How do we make reliability, resiliency, and security part of the fundamental requirements and architecture of future technologies?
10. What does the reliability and security engineering community need to provide as guidelines and/or standards to ensure security assurance in product/technology development?

Update on AI Regulation

(i) Focus areas of AI Regulation

Operational impact

- Risk of unfair bias/discrimination if AI systems are used for profiling or decision-making
- Use of data for model training without permission/control/compensation". Focus on copyrighted materials, personal/private data
- Environmental impact
- Lack of transparency/human oversight

Wider impact on society

- Disruption to jobs and employment shifts
- Misuse of AI by bad actors
- Problem of AI alignment (worsened by rapid acceleration in capabilities)
- Shifting balance of power

(ii) Milestones in AI Regulation

Key changes over the last year:

- Recent milestones in AI regulation
  - Increased focus on 'foundation models' and broader AI safety issues beyond fairness/data – driven initially by UK
  - The US is no longer on the sidelines, and is leading by example with AI regulation for Federal agencies
- Key legislation:
  - **US Executive order** on safe, secure and trustworthy development and use of AI. The US has seized the lead in setting standards for AI responsibility – in particular, for "foundation models". The bulk of the Executive Order is focused on Federal Agencies rather than private companies – but will have wider influence:
    - 150+ actions spanning 50+ federal entities
    - Ambitious deadlines – 69 actions were due within 180 days (by end April 2024)
    - Introduced reporting requirements for foundation models and large-scale compute capacity

  - **European AI Act**. European Parliament approves **final text of AI Act** on 13 March 2024. Trialogues are over, and the text is now final – although still awaiting formal publication. Broad shape of regulation is unchanged, although details have been finessed. Still waiting for clarification on standards:
    - Prohibitions on certain uses of AI
    - Mandatory requirements for "high risk AI systems", "general purpose AI models" and some other narrow applications
    - Key emphasis is on risk/impact assessment and transparency
    - Differing obligations for providers vs deployers
    - Exemptions for R&D, most open source, some "grandfathered" products, law enforcement/defence
    - Some flexibility for products already subject to regulation

  - Enforcement via large fines and new oversight bodies including Office of AI and market surveillance authorities.
  - Staggered implementation from end 2024-2026.

(iii) Summary

- Increased focus on 'foundation models' and broader AI safety issues beyond fairness/data – driven initially by UK
- The US is no longer on the sidelines, and is leading by example with AI regulation for Federal agencies
  - Executive Order (October 2024)
- AI oversight in UK and US is predominantly sector specific, ss. Europe which has taken a broader 'horizontal' approach
  - AI Act (May/June? 2024) – but delayed enforcement.

Responsible AI

Scaling Responsible AI

- AI Responsibility is measured by testing for policy compliance against policies around (for example) solicitation of PII, Hate speech, harassment, etc.
- Current foundation models may have some degree of responsibility built in; however, it may not align with the application users' unique needs.
- The next frontier is to provide users with the tools to build their own models responsibly. Key enablers are – ability to define custom policies, ability to generate high quality policy-specific data for training and testing, ML training and tuning methods that are data-efficient (don't rely on a lot of data)

AI for Communications Networks for AI

- While 6G wireless communication networks orks will include AI capabilities supported by edge computing resources to support delay-sensitive throughput-demanding AI-enabled applications, it is no less true that such complex networks will have to be self-configuration, self-optimization and self-healing, which requires AI mechanisms. This means that, in future networks, AI mechanisms serving both purposes (user applications and communication network operation) will have to be seamlessly embedded in the network architecture from design, instead of as an add-on.
- Dependability of the communication network is expected to significantly improve as embedded AI mechanisms allow higher degrees of self-organization. However, the increased complexity of 5G wireless communication networks presents increased challenges to these AI mechanisms. As an example, in 5G, network slicing makes root cause analysis more difficult, as some anomalies may be caused by complex inter-slice interactions that are difficult to identify.

AI for Sustainability

1. Current Energy Sustainability Numbers:
   - -87% generation from lignite (vs. 2005)
   - 2x solar + wind capacity vs. 2019
   - 48% of power from wind + solar (2M 2024)
   - 4th in wind + solar penetration in the European Union -EU (2023)
   - 2x investment in grids (vs. 2015-18)
   - 11.3% EV market share (vs. 0.4% in 2019
2. There is an inherent limitation in increasing the renewable energy share from the current 48% due to immense storage requirements.
3. AI can help increase this percentage in a cost-efficient manner.
4. AI Intelligence can optimize the Return on Investment (RoI) of renewable energy production and storage investments.
5. The latest advances in Deep Learning from Open AI, Gemini, etc. can be a valuable resource for building the future AI powered energy sustainability solutions.

Reliability and Serviceability

1. Battery Reliability

The reliability, safety, and sustainability of battery is playing an increasingly critical role in the emerging technologies such as next generation communication networks and application devices.

Following are some technical takeaways from the related presentation:

- Battery safety and durability is a highly complicated problem
- Data-driven has demonstrated the strong capability to solve complex system problem
- Predict safety risk
- Classify safety status
- Predict the short circuit resistance (determinant factor)

Following is a partial list of grand opportunities:

- Establish testing methodology for battery safety and durability
  - Work with FM Global
- Establish shared database
  - Work with academic and commercial organizations
- Room for improvement:
  - Develop fast, safe, low-cost testing/characterization
  - Dive deep into fundamental mechanism and develop new physics-based models
  - Deal with multiscale descriptions
  - For ML
    - How to deal with insufficient data or missing data
    - How to deeply interact data-driven and physics
    - How to use ML in and characterization

- How to use LLM in collecting data and feeding it into the database.

2. *Cloud Computing Reliability*

Cloud computing facilitates various service models including communications (e.g., Multiple Access Edge Computing), Healthcare, Industrial IoT (IIoT), Autonomous Ground Vehicles (AGV) among others, in which Reliability is paramount. Some of the noted challenges include:

- Proliferation of edge devices (e.g., IoT/IIoT, Autonomous Vehicles) amplify the attack surface in communication networks and consequently the reliability and security of cloud implementations and resources (e.g., services, applications, user data).
- The rapid growth of cloud computing implementations demands more equipment and power consumption which introduces challenges to environmental sustainability by increasing the carbon footprint produced by computing and data transmission.
- Intense AI workloads and adversarial AI attacks may impact the performance and reliability especially in multi-tenant environments.
- The cloud computing hardware is pretty much based on 3 9s reliability but it is expected to deliver 6 9s reliability if cloud computing is playing an important role of the emerging technology and market such as AI enabled applications and network virtualizations. This will be an essential infrastructure that requires resilient architecture design and AI driven self-healing capabilities. It was excellent to include such a topic in this forum so the engineering and academia communities would look at the software reliability to be an integral part of this complex ecosystem.

Security

Following are some key points from the security-related presentations and discussion:

(i) Security Challenges

5G Security

- 5G exposes a larger attack surface and more potential entry points for malicious actors, due to its increased complexity, heterogeneity, and connectivity.
- 5G supports critical and sensitive applications, such as autonomous vehicles, smart grids, industrial operations, and remote surgery, that require high levels of security and trust.
- 5G involves multiple stakeholders and entities, such as network operators, service providers, device manufacturers, and users, that need to cooperate and coordinate to ensure the security of the network.

Wi-Fi Security

- Wi-Fi 6 is designed to support various use-cases including dense industrial automation environments, smart-homes/offices, transportation, health, and education. However, Wi-Fi 6 operates in both the 2.4 GHz and 5 GHz frequency bands, which are also used by other wireless technologies, such as Bluetooth and Zigbee along with appliances such as microwave ovens and introduce concerns for spectrum sharing and interference.

- Wi-Fi 6 introduces a new version of the Wi-Fi Protected Access security protocol (WPA3) to remediate prior weaknesses but was found to be susceptible to certain attacks which allow an adversary to exploit weaknesses in the handshake and password selection processes (Dragonblood).
- Wi-Fi 6 devices and access points may not support WPA3, or may support different versions or configurations of WPA3, which may introduce interoperability and security issues. Furthermore, legacy devices may still use WPA2 which can be exploited by adversaries to compromise the security of the entire network.

(ii) 5G Threat Domains – Top 10

1. User equipment and edge devices (e.g., UE, IoT devices, gNB/eFemtos/extenders)
2. RAN Signaling
3. 5G Core Signaling
4. Network Slicing
5. Network Peering Functions – Security Edge Protection Proxy (SEPP) ♦ Partner networks
6. Network Exposure Functions (NEF)
7. Network Infrastructure (fronthaul/mid-haul/backhaul)
8. Virtualization / Cloud Infrastructure / MEC (Multi-access Edge Computing)
9. Management and Network Orchestration Applications (MANO/OAM&P/OSS)
10. Software Supply Chain (SBOM)

6G and the Expectation of 5G+ Advancement

The global deployment and the success of 5G are very uneven across the world, ranging from millions of new radio stations of China, South Korea, Japan, Europe, and US to much less coverage in the rest of the world. As the world expands the 5G deployment and applied this new technology to benefits new applications, the demands of speed and latency upgrades are coming in at a modest pace. With the 5G SA commercial network deployment expected in next 2-3 years we will get the first impression of the actual delivery of 5G promises, such as multi Gbps bandwidth and few ms latency for critical applications in combination with edge computing.

There will be an actual assessment of the competence of these commercial networks to support critical applications such as driverless cars, etc. One main challenge will be the ubiquitous coverage required and ~0% drop call rate.

Reliability will be of paramount importance for the continuous service provision along the road networks.

6G along with the promise of AI automated operations, shall address this issue, and introduce additional spectrum at the low bands, to eliminate the blank spots along the highways.

Another question is when 6G planning on reliability needs to start.

ETR-RT24 Open Dialog Wrap-up Summary and Recommendations

Following are highlights from the ETR-RT Open Dialog for the Wrap Up Summary that took place on May 22, 2024:

1. Trustworthiness of Future AI remained a concern for most of the participants.
2. What is the AI platform vs. application concerns and measures?
3. How do we align user (both consumer and business) expectations vs. technology advancement?
4. We must increase the joint efforts by the industry and academia in seamless integration of various technologies for our future.
5. It is a consensus that the Responsive AI must ensure user security for future products out of the gate. We also believe that the security of AI, Hyper Cloud, 5G (even 6G soon) will drive the user security.
6. We strongly recommend that we must promote rigorous and AI driven testing methodologies in both intrusive and non-intrusive, and vertical and horizontal ways to enhance the ETE ecosystem reliability.
7. We are facing a challenge of a five 9s reliability but only one 9 reliability reality for AI today. Recommend industry and academia efforts to drive standardization of AI Reliability Definition, Policies, and Measurements.
8. How do we leverage cryptographic frameworks toward zero-tolerance (or zero-trust) security architecture.
9. The ETR Roundtable recognizes that we need to build/increase future infrastructure/engine to a high degree of reliability for Responsive AI and/or AI driven applications.
10. Recommend Responsive AI and reliable infrastructure to be advanced as a service to enable easy integration and rapid development, as well as promoting awareness and education of such for our next generation talent.

Few Recommendations for IEEE and Next Year's ETR-RT

1. Include Generative AI/Responsible AI as part of future ETR-RTs.
2. Continue to sponsor ETR-RT25 with the expanded scope, including software and critical hardware reliability.

3. Promote awareness of the urgent needs for the reliability of the new emerging technologies, and ETE ecosystem.
4. With the expanded scope and interests, considering extending the conference duration to 2 and half days, including more panel discussions and specific recommendations to industry and academia communities.
5. Recommend IEEE to publish a special paper edition of emerging technology reliability, safety, and security.
6. Recommend the ETR to be a special track to the ICC and GLOBECOM annual conferences.
7. Promote industry investment to the study and research of ETR as a special topic.
8. Ensure reliability, safety, and security to be critical and essential part of emerging technology development for the future.