O-RAN Next Generation Research Group (nGRG)

Contributed Research Report

**Agentic Telco: Threat Landscape for 6G Networks**

**Report ID: RR-2025-0X**

**Contributors:**

**Reliance Jio**

**Rakuten**

**CICT**

**Nokia**

**Digital Catapult**

***Release date: 2025.12.XX***

# **Authors**

**Ravi Sinha, Reliance Jio (Editor)**

**Clifton Fernandes, Anmol Agarwal, Nokia**

**Shihan Bao, CICT**

**Krishna Pramod A, Paromita Chintan Shah, Rakuten**

**Keerthi Thomas, Digital Catapult**

# **Reviewers**

# **Disclaimer**

The content of this document reflects the view of the authors listed above. It does not reflect the views of the O-RAN ALLIANCE as a community. The materials and information included in this document have been prepared or assembled by the above-mentioned authors and are intended for informational purposes only. The above-mentioned authors shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of this document subject to any liability which is mandatory due to applicable law. The information in this document is provided ‘as is,' and no guarantee or warranty is given that the information is fit for any particular purpose*.*

# **Copyright**

The content of this document is provided by the above-mentioned authors. Copying or incorporation into any other work, in part or in full of the document in any form without the prior written permission of the authors is prohibited.

# **Executive summary**

Contents

[**Authors** 2](#_Toc197583551)

[**Reviewers** 2](#_Toc197583552)

[**Disclaimer** 2](#_Toc197583553)

[**Copyright** 2](#_Toc197583554)

[**Executive summary** 2](#_Toc197583555)

[**List of abbreviations** 5](#_Toc197583556)

[**List of figures** 5](#_Toc197583557)

[**List of tables** 5](#_Toc197583558)

[**1** **Introduction** 6](#_Toc197583559)

[**1.1** **Overview** 6](#_Toc197583560)

[**2** **Motivation** 7](#_Toc197583561)

[**3** **Framework for agentic communication security landscape** 8](#_Toc197583562)

[3.1 Agentic Communication in 6G Telco Networks with a baseline architecture. 8](#_Toc197583563)

[3.2 Security Threat Landscape for Agentic Communication 8](#_Toc197583564)

[3.3 Autonomous Defense Mechanisms 8](#_Toc197583565)

[3.4 Trust and Transparency in Agentic Systems 8](#_Toc197583566)

[3.5 Regulatory and Ethical Implications 8](#_Toc197583567)

[3.6 Future Research and Innovations for 6G Agentic Communication 8](#_Toc197583568)

[3.7 Policy and Industry Recommendations 8](#_Toc197583569)

[3.8 Liability for AI-driven Multi Agent System 9](#_Toc197583570)

[3.9 AI agents compliance with telecom security standards 9](#_Toc197583571)

[3.10 Multi Agent System comparison with centralized Security Information 9](#_Toc197583572)

[3.11 Major techniques to improve MAS 9](#_Toc197583573)

[3.11.1 Defensive distillation 9](#_Toc197583574)

[3.11.2 Adversarial training 9](#_Toc197583575)

[3.11.3 Moving target defense (MTD) 9](#_Toc197583576)

[3.12 Secure MPC techniques for MAS nw to analyze encrypted network traffic 9](#_Toc197583577)

[3.13 Methodes to prevent conflict among independently acting agents 9](#_Toc197583578)

[3.14 Differential and individual privacy for MAS 9](#_Toc197583579)

[3.15 General Data Protection methodes for MAS privacy-by-design principles 9](#_Toc197583580)

[3.16 Self-Sovereign identity (SSI) frameworks for MAS 9](#_Toc197583581)

[3.17 Human-in-the-loop (HITL) security mechanisms for MAS 9](#_Toc197583582)

[3.18 NLP interfaces for MAS driven security 9](#_Toc197583583)

[3.19 MAS zero-day attacks in telecom networks. 9](#_Toc197583584)

[3.20 FL approaches for secure, privacy-preserving MAS 9](#_Toc197583585)

[**4** **Research methodology** 10](#_Toc197583586)

[**5** **Use case analysis** 10](#_Toc197583587)

[**5.1** **Exploring use cases and key capabilities** 10](#_Toc197583588)

[**5.2** **Selecting appropriate use cases** 10](#_Toc197583589)

[**5.3** **Use case requirements and gap analysis** 10](#_Toc197583590)

[**6** **Selecting 6G documents for MAS Security** 10](#_Toc197583591)

[**7** **Characteristics of MAS security Landscape for 6G nw** 10](#_Toc197583592)

[**8** **6G capabilities for agentic communication security** 11](#_Toc197583593)

[**9** **Defining criteria for gap analysis** 11](#_Toc197583594)

[**10** **Considerations** 12](#_Toc197583595)

[**11** **Conclusion** 12](#_Toc197583596)

[**12** **References** 12](#_Toc197583597)

[**13** **Revision History:** 12](#_Toc197583598)

# **List of abbreviations**

|  |  |
| --- | --- |
| AGV | Autmatic Guided Vehicle |
| AI | Artificial Intelligence |
| AIC | AI Integrated Communications |
| DT | Digital Twins |
| E2E | End to End |
| MIMO | Multiple Input Multiple Output |
| mMTC | massive Machine Type Communications |
| MNO | Mobile Network Operator |
| nGRG | Next Generation Research Group |
| NTN | Non-Terrestrial Network |
| O-CU | O-RAN Central Unit |
| O-DU | O-RAN Distributed Unit |
| O-RU | O-RAN Radio Unit |
| RAN | Radio Access Network |
| RIS | Reconfigurable Intelligent Surface |
| RF | Radio Frequency |
| RS | Research Stream |
| SDGs | Sustainable Development Goals |
| Sub6 | Sub 6GHz |
| SUC | Summrized Use Case |

# **List of figures**

[Figure 1 8](https://dell-my.sharepoint.com/personal/mohammad_alavirad_dell_com/Documents/Advanced%20Wireless%20Technology/Standard/O-RAN/nGRG/RS01%20-%20Requirements/01%20-%20mmWaveAntennaExtension/05%20-%20Research%20Report/33%20-%20EC%20review/nGRG-RR-2024-04-UC_analysis_mmWave_antenna_dist.docx#_Toc171422760)

# **List of tables**

# **Introduction**

# **Overview**

Today’s 5G networks are under attack—and legacy defenses are losing the war. They’re bigger, more complex, and more critical than ever, but the old ways of securing them just aren’t cutting it anymore. Since AI/ML framework is growing with the advancement of Telco Networks faster than the expectations, it is important to consider and evaluate, how the security framework will be effective in the ever-growing era of Agentic Communication.

An AI Agent is a software system that understands its environment, processes information, and acts independently to achieve goals. It uses tools like machine learning and decision-making to handle tasks effectively with little human help. AI agent interacts with each other and humans using natural and intuitive language. It emphasizes clarity, autonomy, and effectiveness in achieving tasks through conversation.

When Agentic Communication comes in. It is a network that doesn’t just respond to threats but thinks ahead—like a digital immune system with built-in AI brains. These systems don’t wait for human commands. They detect, adapt, and fight back in real time, all on their own. Along with all these benefits, it also brings lots of challenges. Autonomous Decision Risks, Adversarial AI Exploitation, Distributed Trust Management, Emergent Behavior Complexities and Protocol-Level Vulnerabilities are some of the new sets of challenges come with Agentic communication framework in Telco infrastructure.

This Research Report (RR) will dissect every layer of Agentic Communication security, establishing an intelligent framework to future-proof autonomous networks against evolving AI/ML threats in telecom infrastructure.

A diagram of a car

Description automatically generated

# **Motivation**

In an era where connectivity is the backbone of global progress, telecommunications networks stand as the lifeline of modern communication. Yet, with growing digital transformation comes escalating cyber threats—unauthorized intrusions, data breaches, and sophisticated attacks that jeopardize trust and operational integrity. Agentic communications further complicate the security challenges in Telco domains. Here are some of the challenges, which equally considered the opportunities and motivation for this RID

* Increasing Reliance on Autonomous Systems-- If agents are compromised, entire telecom infrastructures (e.g., network slicing with highly distributed AI enabled Edge Infrastructure) become vulnerable.
* Without secure agent communication, attackers could- Redirect traffic (BGP hijacking via malicious agents).Disrupt critical services (e.g., emergency communications).Steal sensitive data (e.g., user metadata, call logs).
* Regulatory & Compliance Requirement for the Multi Agent System for Telecommunication Network.
* Need for Real-Time, Adaptive Security
* Trust in Multi-Agent Collaboration

Privacy Preservation in Distributed AI

# **Framework for agentic communication security landscape**

These are the basic building blocks of our objective to this research report:

## Agentic Communication in 6G Telco Networks with a baseline architecture.

A screenshot of a computer

Description automatically generated

## Security Threat Landscape for Agentic Communication

Agentic communication, where autonomous AI agents interact seamlessly is reshaping digital communications. But with this rapid evolution comes new security risks that industry can’t afford to overlook. As AI-driven interactions become more common, Telco Industry must stay ahead of emerging threats to ensure these systems remain reliable, safe, and protected from exploitation. Here is the threat landscape in the world of Agentic communication

* **Ensuring Data Security and Trust –** Memory Injection, Data Leaks, Privacy Violations, Spoofing and Manipulation are some of the major factors falls under this category.
  + **Data Leaks:** Vulnerabilities in agent software or communication channels can lead to the exposure of sensitive data.
  + **Memory Injection:** Adversaries can inject misleading information into an agent's memory, influencing its future behavior
  + **Privacy Violations:** Agents collecting and processing personal data raise concerns about consent, usage, and potential misuse.
  + **Spoofing and Manipulation:** Malicious actors might try to alter data exchanged between agents or with users**.**
* **Integrity and Trust:**
  + **Code Tampering:** Modifying agent code to introduce malicious functionalities.
  + **Rogue Agents:** Introducing malicious agents into a multi-agent system.
  + **Compromised Decision-Making:** Manipulating agent reasoning to make harmful choices.
  + **Trust Establishment:** The inherent difficulty in trusting autonomous systems without clear accountability.
  + **Cross-System Vulnerabilities** – AI agents interacting with multiple enterprise systems may introduce security risks through lateral movement.
* **Authentication and Authorization:**
  + **Impersonation:** Attackers could attempt to impersonate legitimate agents or users.
  + **Unauthorized Control:** Gaining control over an agent's actions or capabilities.
  + **Credential Theft:** Compromising agent credentials to access tools, data, or systems under a false identity.
  + **Weak Access Controls:** Insufficiently granular control over what agents can access and do.
  + **Overprivileged AI Agents** – AI systems often inherit permissions from users, potentially accessing restricted data and bypassing security controls.
* **Availability and Resilience:**
  + **Denial of Service (DoS/DDoS):** Overwhelming agent communication infrastructure to disrupt operations.
  + **Targeted Attacks:** Focusing on critical agent components to cause system failure.
  + **Single Points of Failure:** Lack of redundancy in agentic systems.
* **Novel Threats Specific to Agentic Systems:**
  + **Adversarial Attacks on Agent Intelligence:** Crafting inputs to mislead agent perception or reasoning (e.g., manipulating sensor data in autonomous vehicles).
  + **Social Engineering of Agents:** Exploiting agent vulnerabilities to influence their behavior through deceptive prompts or interactions.
  + **Emergent Malicious Behavior:** Unexpected harmful actions arising from complex multi-agent interactions.
  + **Lack of Visibility and Monitoring** – Organizations struggle to track AI usage, increasing the risk of unauthorized data exposure.
  + **Control and Governance Issues:** The challenge of maintaining oversight and accountability as agent autonomy increases.
  + **Hallucination and Misinformation** – AI-generated responses may contain incorrect or misleading information, impacting security and compliance.

## Autonomous Defense Mechanisms

## Trust and Transparency in Agentic Systems

## Regulatory and Ethical Implications

## Future Research and Innovations for 6G Agentic Communication

## Policy and Industry Recommendations

## Liability for AI-driven Multi Agent System

## AI agents compliance with telecom security standards

## Multi Agent System comparison with centralized Security Information

## Major techniques to improve MAS Security

The path to truly secure Multi-Agent Systems is paved with the mastery of these major techniques

## Defensive distillation

In the high-stakes arena of telecom, the vulnerability of AI agents is a direct threat to critical telecom infrastructure. Defensive distillation offers a powerful solution as an armor against adversarial attacks that ensures the reliability and integrity of our increasingly autonomous systems. While implementation demands rigor, the payoff is an unyielding network built on trustworthy AI, which is essential for safeguarding the future of communication. Further research and experimentation are warranted to optimize the application of defensive distillation within the specific context of various telecom AI agentic communication scenarios.

Model hardening techniques, aims to improve the resilience of machine learning models against adversarial attacks, under Defensive distillation framework . It involves a two-stage training process:

* **Training the Original Model:** A standard machine learning model is trained on the original dataset. This is called the Teacher model.
* **Training the Distilled Model:** A new, smaller student model is trained using the softened probability outputs of the larger, pre-trained teacher model. The temperature parameter (T) is used to soften these probabilities, making the model less confident in its predictions and smoothing the decision boundaries. Training with these softened probabilities helps the student model learn the relationships between different classes in a smoother way.

There are several key benefits Implementing defensive distillation in telecom AI agentic communication framework:

* **Model Compression for Distributed Telco Networks:** The distilled student model is often smaller than the original teacher model, potentially leading to reduced computational overhead and faster inference times, which are critical for real-time telecom applications.
* **Improved Model Generalization:** The softened probability targets during distillation can lead to better generalization and more stable predictions on unseen data, including potentially noisy or slightly manipulated inputs.
* **Increased Robustness Against Adversarial Attacks:** Distilled models are generally more resilient to small, imperceptible perturbations in the input data that can fool standard models. This is crucial for preventing evasion attacks on critical telecom AI functions.
* **Enhanced Trust and Reliability:** For AI enabled Telco infrastructure and making AI agents more resistant to manipulation, defensive distillation contributes to greater trust and reliability in their decisions and actions.

Despite its promise, the integration of defensive distillation into telecom AI agentic communication is fraught with tangible challenges that cannot be ignored.:

* **High Performance Computational (HPC) Cost:** The two-stage training process can be computationally expensive, especially for large and complex models used in telecom applications. Deployment of Small Language models specific to the telecom domain may reduce the deployment cost drastically
* **Hyperparameter Tuning:** The temperature parameter (T) in the distillation process needs careful tuning to achieve the right balance between robustness and accuracy. Suboptimal values can lead to a significant drop in performance on clean data.
* **Sensitivity to Attack Types:** While effective against many types of adversarial attacks, defensive distillation might not provide complete protection against all sophisticated attack strategies. Combining it with other defense mechanisms is often necessary and adds extra cost for the telco network.
* **Model Accuracy impact:** In some cases, defensive distillation can lead to a slight decrease in accuracy on clean data, requiring a careful trade-off between robustness and performance.
* **Complexity of Multi-Agent Systems:** Applying distillation across a network of interacting agents requires careful consideration of how the robustness of individual agents contributes to the overall system security.

## Adversarial training

## Moving target defense (MTD)

## Secure MPC techniques for MAS nw to analyze encrypted network traffic

## Methodes to prevent conflict among independently acting agents

## Differential and individual privacy for MAS

## General Data Protection methodes for MAS privacy-by-design principles

## Self-Sovereign identity (SSI) frameworks for MAS

## Human-in-the-loop (HITL) security mechanisms for MAS

## NLP interfaces for MAS driven security

## MAS zero-day attacks in telecom networks.

## FL approaches for secure, privacy-preserving MAS

# **Research methodology**

This chapter will briefly describe the methodology used to conduct this research work.

# **Use case analysis**

# **Exploring use cases and key capabilities**

# **Selecting appropriate use cases**

# **Use case requirements and gap analysis**

In this chapter we will delve into the specific needs of our chosen 6G use cases and explore the potential integration of Security Landscape for Agentic Communication.

# **Selecting 6G documents for MAS Security**

To investigate the use cases, the following documents have been selected:

# **Characteristics of MAS security Landscape for 6G nw**

# **6G capabilities for agentic communication security**

# **Defining criteria for gap analysis**

# **Considerations**

In this chapter, our focus shifts to

# **Conclusion**

This research aims to bridge the gap between autonomous AI agents and next-gen telco security, offering actionable insights to mitigate risks while harnessing the power of agentic communication for a safer, smarter digital future

# **References**

Multi Agent System: Survey: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8352646

Artificial Intelligence Risk Management Framework (AI RMF 1.0): //efaidnbmnnnibpcajpcglclefindmkaj/https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf (RMF)

# **Revision History:**

|  |  |  |  |
| --- | --- | --- | --- |
| Version number | Author | Date of revision | Description of changes |
| 1.0.0 | Ravi Sinha | 8th May 2025 | Initial Draft |
| 1.0.1 | Ravi Sinha | 8th May 2025 | Rearrangement of ToC |
| 1.0.2 | Ravi Sinha | 14th May 2025 | 3.2 Security Threat Landscape for Agentic Communication added |
| 1.0.3 | Ravi Sinha | 15th May 2025 | 13.1 Major techniques to improve MAS Security  13.1.1Defensive distillation  added |