# AICT 2023

The Nineteenth Advanced International Conference on Telecommunications

ISBN: 978-1-68558-068-1

June 26 - 30, 2023

Nice, France

**AICT 2023 Editors**

Eugen Borcoci, University Politehnica of Bucharest, Romania

# AICT 2023

# Forward

The Nineteenth Advanced International Conference on Telecommunications (AICT 2023), held between June 26th and June 30th, 2023, continued a series of events covering a variety of challenging telecommunication topics ranging from background fields like signals, traffic, coding, communication basics up to large communication systems and networks, fixed, mobile, and integrated, etc. Applications, services, system, and network management issues also received significant attention.

The spectrum of 21st Century telecommunications is marked by the arrival of new business models, new platforms, new architectures, and new customer profiles. Next generation networks, IP multimedia systems, IPTV, and converging network and services are new telecommunications paradigms. Technology achievements in terms of co-existence of IPv4 and IPv6, multiple access technologies, IP-MPLS network design driven methods, multicast and high speed require innovative approaches to design and develop large scale telecommunications networks.

Mobile and wireless communications add profit to a large spectrum of technologies and services. We witness the evolution 2G, 2.5G, 3G and beyond, personal communications, cellular and ad hoc networks, as well as multimedia communications.

Web Services add a new dimension to telecommunications, where aspects of speed, security, trust, performance, resilience, and robustness are particularly salient. This requires new service delivery platforms, intelligent network theory, new telecommunications software tools, new communications protocols, and standards.

We are witnessing many technological paradigm shifts imposed by the complexity induced by the notions of fully shared resources, cooperative work, and resource availability. P2P, GRID, Clusters, Web Services, Delay Tolerant Networks, Service/Resource identification and localization illustrate aspects where some components and/or services expose features that are neither stable nor fully guaranteed. Examples of technologies exposing similar behavior are WiFi, WiMax, WideBand, UWB, ZigBee, MBWA and others.

Management aspects related to autonomic and adaptive management includes the entire arsenal of self-ilities. Autonomic Computing, On-Demand Networks and Utility Computing together with Adaptive Management and Self-Management Applications collocating with classical networks management represent other categories of behavior dealing with the paradigm of partial and intermittent resources.

We take here the opportunity to warmly thank all the members of the AICT 2023 technical program committee, as well as all the reviewers. The creation of such a high-quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to AICT 2023. We truly believe that, thanks to all these efforts, the final conference program consisted of top-quality contributions. We also thank the members of the AICT 2023 organizing committee for their help in handling the logistics of this event.

We hope that AICT 2023 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the field telecommunications.

**AICT 2023 Chairs**

**AICT 2023 Steering Committee**

Dragana Krstic, University of Niš, Serbia
Kevin Daimi, University of Detroit Mercy, USA
Gautam Srivastava, Brandon University, Canada
Sergei Semenov, HiSilicon, Sweden

**AICT 2023 Publicity Chairs**

Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain
José Miguel Jiménez, Universitat Politecnica de Valencia, Spain

# AICT 2023
# Committee

**AICT 2023 Steering Committee**

Dragana Krstic, University of Niš, Serbia
Kevin Daimi, University of Detroit Mercy, USA
Gautam Srivastava, Brandon University, Canada
Sergei Semenov, HiSilicon, Sweden

**AICT 2023 Publicity Chairs**

Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain
José Miguel Jiménez, Universitat Politecnica de Valencia, Spain

**AICT 2023 Technical Program Committee**

Ghulam Abbas, GIK Institute, Pakistan
Iwan Adhicandra, University of Sydney, Australia
Abdullah Al-Alaj, Virginia Wesleyan University, USA
Abdulazaz Albalawi (Aziz), University of California, Santa Cruz, USA
Michele Albano, Aalborg University, Denmark
Nicolae Dumitru Alexandru, "Gheorghe Asachi" Technical University of Iasi, Romania
Marco Aurélio Spohn, Federal University of Fronteira Sul, Brazil
Ilija Basicevic, University of Novi Sad, Serbia
Oussama Bazzi, Lebanese University, Lebanon
Carlos Becker Westphall, Federal University of Santa Catarina, Brazil
Sara Behjatjamal, Altinbas University, Turkey
Stefano Berretti, University of Firenze, Italy
Robert Bestak, Czech Technical University in Prague, Czech Republic
Razvan Bocu, Transilvania University of Brasov, Romania
Antonella Bogoni, Scuola Superiore Sant'Anna-TeCIP Institute, Italy
Eugen Borcoci, University "Politehnica"of Bucharest (UPB), Romania
Christos Bouras, University of Patras, Greece
An Braeken, Vrije Universiteit Brussel, Belgium
Lubomir Brancík, Brno University of Technology, Czech Republic
Peter Brida, University of Zilina, Slovakia
Ivana Bridova, University of Zilina, Slovakia
Nicola Calabretta, Eindhoven University of Technology, Netherlands
Vítor Carvalho, 2Ai-EST-IPCA / Algoritmi Research Center - Minho University, Portugal
Fernando Cerdan, Polytechnic University of Cartagena, Spain
Adil Chekati, University of Tunis El Manar, Tunisia
Chen Chen, Chongqing University, China
Enrique Chirivella Pérez, Universitat de Valencia, Spain
Tanveer Choudhury, Federation University Australia, Australia
Chi-Wai Chow, National Chiao Tung University, Taiwan
Giampiero Contestabile, Scuola Superiore Sant'Anna, Pisa, Italy

Vladica Sark, IHP - Leibniz-Institut für innovative Mikroelektronik, Frankfurt (Oder), Germany
Vincent Savaux, b<>com, Rennes, France
Antonio Luiz Schalata Pacheco, Federal University of Santa Catarina, Brazil
Motoyoshi Sekiya, Fujitsu, Japan
Sergei Semenov, HiSilicon Lund Research Center, Sweden
Ana Serrano Tellería, University of Castilla La Mancha, Spain
A. F. M. Shahen Shah, Yildiz Technical University, Istanbul, Turkey
Yingjie Shao, The Chinese University of Hong Kong, Hong Kong
Shuaib Siddiqui, Fundacio i2CAT, Barcelona, Spain
Zdenek Smekal, Brno University of Technology, Czech Republic
Antonio Sorrentino, Università Parthenope, Napoli, Italy
Gautam Srivastava, Brandon University, Canada
Kostas Stamos, University of Patras, Greece
Horia Stefanescu, Orange, Romania
Oscar Tamburis, Università "Federico II" Napoli, Italy
Mahdi Tousizadeh, University of Malaya, Malaysia
Angelo Trotta, University of Bologna, Italy
Thrasyvoulos Tsiatsos, Aristotle University of Thessaloniki, Greece
Sezer Ulukaya, Trakya University, Turkey
Rob van der Mei, Centre for Mathematics and Computer Science (CWI), Netherlands
Bernd E. Wolfinger, University of Hamburg, Germany
Xuwei Xue, Beijing University of Posts and Communications, China
Ramin Yahyapour, Georg-August-Universitaet Goettingen/GWDG, Germany
Hui Yang, Beijing University of Posts and Telecommunications (BUPT), China
Mehmet Akif Yazici, Informatics Institute - Istanbul Technical University, Turkey
Chien-Hung Yeh, Feng Chia University, Taiwan
Shaohua Yu, Northwestern Polytechnical University, China
Mariusz Zal, Poznan University of Technology, Poland
Francesco Zampognaro, University of Rome "Tor Vergata", Italy
Jiao Zhang, Purple Mountain Laboratories, China
Stefania Zinno, University of Naples Federico II, Italy
Mario Zitelli, Sapienza Univ. di Roma, Italy
Piotr Zwierzykowski, Poznan University of Technology, Poland

**Copyright Information**

**Table of Contents**

# On Techno-Economic Optimization of Non-Public Networks for Industrial 5G Applications

Jyrki Penttinen
Syniverse Technologies
Atlanta, GA, USA
email: jyrki.penttinen@outlook.com

Jari Collin
Telia Company
Helsinki, Finland
email: jari.collin@telia.com

Jarkko Pellikka
Nokia
Helsinki, Finland
email: jarkko.pellikka@nokia.com

***Abstract*—Non-Public Networks (NPN), or Private Networks, are gaining traction in commercial deployments as they provide benefits to many verticals. The technical base of the NPN is being developed in 3rd Generation Partnership Project (3GPP), and the currently available models can already be deployed providing ecosystem multiple options to set up the services basing on various mobile communication generations. This paper discusses the current state of the art of the standardized NPN solutions focusing on Fifth Generation of mobile communications (5G), and evaluates their applicability to industrial applications. This paper also discusses the ecosystem needs and respective gaps in the models considering selected industrial use cases. Through available references on experiences and deployment scenarios, this paper evaluates some of the key aspects that can impact NPN model selection related to Industrial IoT scenarios, and proposes ways to assess NPN techno-economic aspects.**

***Keywords—NPN; private network; 5G deployment modeling; NPN optimization; network planning***

## I. INTRODUCTION

A Non-Public Network (NPN), referred also to as Private Network, provides network services in isolated environment. It can base on cellular networks or other wireless technologies, and it does not depend on numbering of regulated Public Land Mobile Networks (PLMN). 5G NPN refers to a 3GPP cellular system to deliver its capabilities for NPN use cases, such as businesses and municipalities. The 5G NPN can reside partially or completely in physical premises of an organization using it, e.g., within a factory or campus area, so that an external entity separate from a Mobile Network Operator (MNO) assumes the responsibilities of the isolated part offering its services to a limited group. NPN does not typically allow inbound roamers, although the NPN users may have roaming capabilities to use other PLMNs. The benefits of NPN deployment include the possibility to control the Quality of Service (QoS) and protection by isolation. NPN service can include voice connectivity in defined geographical area, or it can focus on Internet of Things (IoT) and respective industrial applications; today, there are many test projects and commercial setups involving industrial devices [1].

This paper explores the applicability of NPNs for Industrial 5G applications. Industrial 5G refers to Industrial Internet of Things (IIoT) with real-time data of networked sensors, assets, objects, and people, 5G network and edge computing enabling ultra-reliable, low-latency, and high-bandwidth communication.

The presented model is aimed to provide means to evaluate the feasibility of available private network deployment models considering key requirements of use cases of interest. The model supports high-level techno-economic assessment, although its limitation is that it considers rather generic parameters. Along with practical experiences, the model and its use can be developed further to better reflect the realities.

This paper introduces the standardization of private network landscape in Section II, and summarizes the key architectures of the available 3GPP models in Section III. Section IV discusses private network deployment models including 3GPP and other relevant options highlighting their key aspects. Section V summarizes the most important pros and cons of the presented private network variants, comparing their suitability in different scenarios. Section VI presents how techno-economic modeling can be applied in the selection of the most feasible variants, and Section VII summarizes the findings.

## II. STANDARDIZATION

Private networks can be formed by a variety of technologies. To cope with the increased interest and demand, 3GPP has formed a set of standardized solutions based on mobile network technology, and industry bodies are considering guidelines for the actual deployment.

### A. 3GPP

The 3rd Generation Partnership Project (3GPP) has designed 4G and 5G NPN specifications in Release 16 providing enablers for also Industrial 5G IoT. The areas include Time Sensitive Networking (TSN), NPN, and Local Area Network (LAN) type services. Release 17 and beyond evolve these aspects further. The 3GPP defines NPN in the Technical Specifications TS 23.251 (architecture and functional description of network sharing), TS 22.104 (service requirements for cyber-physical control applications in vertical domains), and TS 23.501 (5G System architecture).

### B. Industry bodies

The 5G-ACIA (5G Alliance for Connected Industries and Automation) summarizes industrial IoT deployment scenarios for 5G NPNs basing on 3GPP-defined 5G NPN [2]. It presents deployment models to complement their architectural design.

Also, GSMA (GSM Association) considers NPN, and their guidelines provide overview to deploy 5G industry campus NPN by 3GPP definition, which is one of the key 5G concepts to support “to business” models (2B) [3].

 ISBN: 978-1-68558-068-1 

## III. 5G NPN STANDARD ARCHITECTURES

As per the 3GPP Release 16 definitions, an NPN enables deployment of 5G System (5GS) for private use. The NPN can be deployed as a Stand-alone Non-Public Network (SNPN) or Public Network Integrated NPN (PNI-NPN). An NPN operator manages SNPN without relying on the functions of a PLMN, whereas PNI-NPN deployment depends on those [4].

Figure 1 depicts NPN variants as interpreted from [3], [2].

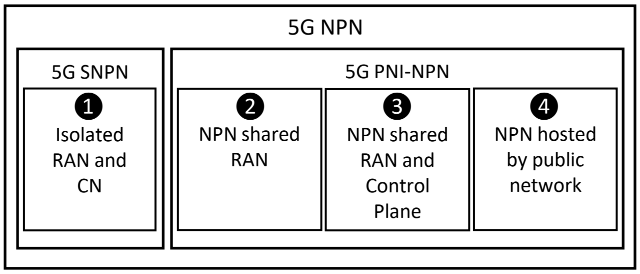


Figure 1 3GPP and 5G-ACIA NPN variants.

### A. SNPN

The SNPN uses combined PLMN Identifier (PLMN ID) and Network Identifier (NID). 5G User Equipment (UE) supporting SNPN can attach to it based on 5G Subscriber's Permanent Identifier (SUPI) and credentials. The Radio Access Network (RAN) of SNPN broadcasts the combined PLMN ID and NID in the System Broadcast Information and supports network selection and re-selection, load and access control, and barring. The NIDs can be self-assigned individually to the SNPN NIDs upon its deployment. The active NIDs may not be unique, but they use different numbering space than the other scenario, coordinated NID-assignment, that can have either 1) globally unique NID-assignment independent of the respective PLMN ID; or 2) globally unique NID/PLMN ID combination.

### B. PNI-NPN

The PNI-NPN uses PLMN ID whereas Closed Access Group Identity (CAG ID) indicates the CAG -enabled 5G radio cells. Within a PLMN, a CAG cell can broadcast one or more CAG IDs, in which case PLMN ID is the base for the network selection and reselection whereas the network uses CAG ID for the cell selection and re-selection, as well as for the control for letting only CAG-enabled UEs access the network.

### C. Implementation aspects

As depicted in Figure 1, these two options result in practical scenarios of isolated deployment of standalone non-public network (1) or non-public network in conjunction with public networks. The latter breaks down into three scenarios including the Industrial and IoT environment: Shared radio access network (2); Shared radio access network and control plane (3); and NPN hosted by the public network (4) [2].

- SNPN: the NPN is separated from the public network and all network functions reside inside the organization's premises. The possible communication between the NPN and the public network takes place via a firewall, e.g., through Non-3GPP Interworking Function (N3IWF).
- Shared radio access network: the NPN and public network share part of the radio access network as per the 3GPP TS 23.251. The communication stays within NPN.
- Shared radio access network and control plane: the NPN and the public network share the RAN for the defined premises while the public network does the network control tasks, the NPN traffic remaining within the premises. Network slicing or 3GPP Access Point Name (APN) can realize this case.
- PLMN-hosted NPN: the enterprise is served by a Network Slice (NS).

## IV. NPN DEPLOYMENT MODELS

### A. Standalone NPN

Figure 2 depicts the principle of SNPN as interpreted from [2], [5]. In this scenario, the User Plane Function (UPF) works as a data router to connect Multi-access Edge Cloud (MEC) and possible LAN. The RAN manages the connectivity of 5G gNB (next generation Node B) and UE of the SNPN users on licensed or unlicensed band. 5G Core (5GC) houses the NFs including UDM (Unified Data Management for user credentials).

In this model, the Network Functions (NF) reside within the operational area of related entity, such as factory, the SNPN being an isolated network from the PLMN. This allows communication between the PLMN and NPN through an optional firewall which isolates the NPN so that Operational Technology (OT) company can operate the NPN and its services, including the NPN IDs to have additional PLMN services in the NPN coverage area, NPN-subscribers to roam the public networks, and public networks' subscribers to roam the NPN depending on roaming agreement. NPN users may also have dual subscription for the PLMN use.

An example of the NPN 5G deployment on IIoT scenarios is the interconnection with TSN as per 3GPP TS 24.519. TSN is a set of new open standards that provide deterministic, reliable, high-bandwidth, low-latency communication [6].

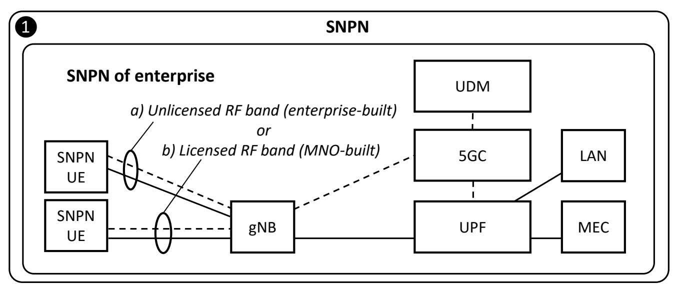


Figure 2 Principle of a SNPN.

The standalone private network can be built in various ways: by dedicated Service Level Assurance (SLA), local PLMN, PLMN by dedicated proportion of operator's licensed spectrum, or by SNPN using unlicensed or private spectrum.

Each deployment model has their pros and cons. As an example, licensed spectrum is one of the most expensive single items in the commercial network. Feasible way to set up this type of network is to construct mm-Wave radio access points in a limited enterprise area and virtualized cloud core functions in nearby edge, one option being a broker managing the NPN [7].

### B. Shared RAN

Shared RAN involves NPN and PLMN with certain part of its RAN for joint use whereas other network functions remain separated. Figure 3 depicts the principle of this model showing the connectivity of the NPN RAN to PLMN core while the own core network of the NPN is isolated from the external world as interpreted from [2], [3], [8]. In this deployment model, NPN traffic stays internal and within the logical, defined area, such as factory premises.

The 3GPP TS 23.251 details the network sharing model in its architectural and functional description and scenarios for network sharing usable also in NPN environment, which are Gateway Core Network (GWCN) and Multi-Operator Core Network (MOCN) [9].

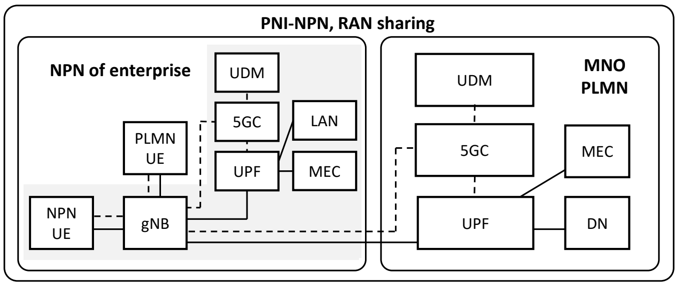


Figure 3 Shared RAN NPN deployment. The grey area indicates the private network slice that forms the NPN.

## C. Shared RAN and Control Plane

As depicted in Figure 4, interpreted from [2], [3], [8], in this deployment, both NPN and PLMN share the RAN within defined business area premises and PLMN takes care of the control plane so that the internal NPN traffic stays always within the logical network related to the business. Network Slicing is one way to set up this scenario as it creates logically independent networks within a shared physical infrastructure. The isolation of the private network portion is possible by using unique NS identifiers.

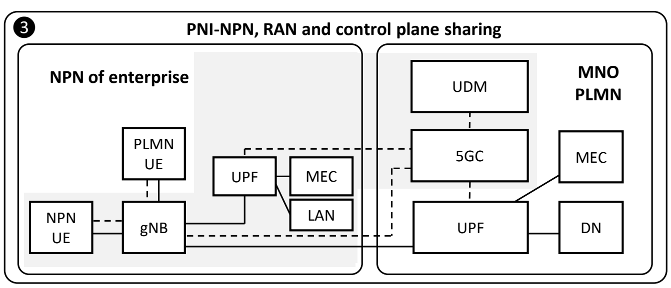


Figure 4 Deployment for shared RAN with control plane. The grey area indicates the private network slice that forms the NPN.

Using APN, as defined by 3GPP, is another way to implement this scenario. In this case, the APN indicates the target network with opportunity to differentiate traffic.

In the shared RAN scenario with shared control plane, the PLMN hosts the NPN so that the devices are a subset of PLMN subscribers. This arrangement eases the contractual aspects of PLMN and NPN operators, and the NPN devices can connect, apart from the NPN itself, also to the PLMN services and roaming. The NPN services may connect to PLMN services, which requires optional interface between the NPN and PLMN services, so NPN devices can connect to NPN services via the PLMN if the device is located outside of the NPN coverage and still within the PLMN. Logically, if the NPN devices can access the PLMN services, this interface is not needed.

## D. PLMN-Hosted NPN

In this case, as depicted in Figure 5 and interpreted from [2], [3], [8], thanks to the network virtualization and cloudification, both the PLMN and NPN traffic are external to the business area. This means that these traffic flows are served by different networks, and the NPN subscribers are in fact public network subscribers. The NPN-PLMN roaming implementation is straightforward as the traffic routes via the PLMN.

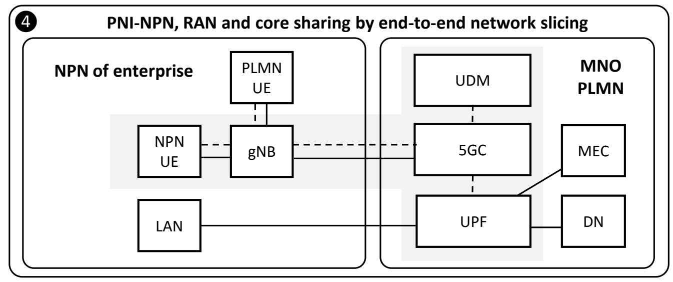


Figure 5 PLMN-hosted NPN. The grey area indicates the private network slice that forms the NPN.

## E. Private network on Network Slice

Although the differentiation between certain user types is possible in 4G networks, it is limited to techniques, such as the isolation of services in a common infrastructure; the means for this include APN Routing, MOCN, and Dedicated Core Network (DECOR) [10]. Built upon Service Based Architecture (SBA), which enables the use of common hardware that executes the Network Functions as instances, 5G has been designed to support also network slicing with respective QoS assurance. In this manner, the Network Slice Provider (NSP) can offer suitable characteristics within their different NSs fulfilling variety of different requirements for their subscribed verticals, also in NPN environment. NS provides means to differentiate the network resources and performance figures that can create also new business models. As an example, the operator can offer their customers gold, silver, and bronze categories, each having their personalized NS price and QoS levels [11]. The NSP can be either MNO or 3rd party. Technically, the NSP could be also enterprise taken their skillset suffices to manage slices.

In NS-based NPN, it is important to adequately interpret the end-user requirements. GSMA provides guidelines for NS setup based on requirements [12], and clarifies how the requirements can be captured from the vertical field [11].

NSs are not yet used widely in commercial Standalone (SA) 5G networks, despite of the forecast indicating about 25% use base in 5G by 2025 [13], [14]. Furthermore, the optimal NS functionality in practice, especially in the end-to-end scenarios, might require still further development to cope with the impacts of real-world non-idealities in synchronization, near real-time orchestration, and overall management of the slices.

## F. Open RAN as a Private Network

Open RAN refers to the overall movement of the telecom industry to disaggregate hardware and software to create respective open interfaces in between [15]. O-RAN Alliance publishes RAN specifications, releases open software for the


ISBN: 978-1-68558-068-1

RAN, and supports O-RAN Alliance members in integration and implementation testing. O-RAN Alliance works on open, interoperable interfaces, RAN virtualization, and big data enabled RAN intelligence [15], [16]. Open RAN may offer feasible possibility to form small-scale, isolated shared RAN networks also in a form of NPN. The concept is still evolving, though, and may not provide optimal techno-economic solutions soon; nevertheless, as the technology matures, Open RAN may offer competitive 4G and 5G NPN variants.

### *G. Legacy network as a service*

Increasing number of incumbent MNOs have switched off their 2G and/or 3G legacy mobile communications networks, or are aiming to sunset them soon. Nevertheless, there will remain scenarios involving IoT and voice services via legacy systems especially in developing markets. According to GSMA statistics, the 2G and 3G systems represent combined still more than 20% of the global footprint in 2025 [14]. This can be a niche opportunity to manage part of this lessening infrastructure on the remaining spectrum, maintaining a minimum feasible infrastructure for IoT and other 2G/3G services via private networks for verticals needing only low capacity.

### *H. Local and Fixed Wireless Access (FWA)*

The LTE-WLAN Aggregation (LWA) as per 3GPP Release 13 allows mobile device configuring on simultaneous LTE and Wi-Fi links [17]. 5G can work also in parallel with non-3GPP accesses to comply with a light-weight private networks' need. 3GPP Release 15 defines Non-3GPP Interworking Function (N3IWF) allowing Wi-Fi access points delivered via 5G infrastructure as Wi-Fi hotspots or Fixed Wireless Access (FWA). The Release 16 defines further Residential Gateways (RG) to interconnect end-users' devices via trusted access points. These gateways complement the N3IWF.

FWA can provide solution to variety of use cases, such as tethering and Mobile Broadband (MBB), best effort FWA, and speed-based QoS [18]. Reflecting these use cases, the FWA could serve as a small-scale home office solution, too, with quick and economic deployment without need for optic fiber.

### *I. Non-3GPP wireless private network*

Technically, it is possible to set up a simple private network using any wireless access beyond the 3GPP specifications on unlicensed, shared spectrum. An example of this is a Wi-Fi hotspot network using common applications within the network providing Over the Top (OTT) voice and messaging. Also, Low Power Wide Area Networks (LPWAN) can provide a feasible communications channel to many IIoT use cases.

## V. CONSIDERATIONS OF SCENARIOS

The feasibility of each NPN deployment depends on the foreseen use cases. As an example, the roaming requirements impact the needed coverage and mobility of the end-users of the NPN. This is one of the examples that can dictate the selection between Standalone and PLMN-assisted scenarios.

### *A. SNSP deployment*

The main aspects of the previously presented deployment models for SNPN scenarios are summarized in the following list, with statements on their suitability for selected use cases.

<u>5G SNPN</u>. OT operates the NPN and its services behind a firewall independently. Provides good security isolation, e.g., to IIoT applications as data is not exposed externally. Optional PLMN interconnectivity via firewall. The operation and management of SNPN requires sufficient skillset from OT company. Provides the opportunity to build a very secure environment, but can be more expensive than only partially owned, or completely outsourced network.

<u>5G SNPN with SLA</u>. The agreed SLA level impacts the business case; the higher the SLA, more costly the CAPEX and OPEX due to, e.g., active-active network mode and reliability of, e.g., 99.999% as per 3GPP Rel 15 URLLC performance, or up to 99.9999% as per Release 16 performance for, e.g., TSN interconnectivity to serve critical IIoT applications.

<u>PLMN with local infrastructure</u>. This is a special case of PLMN that isolates part of the infrastructure and spectrum to a sole use of private network devices in limited geographical area. It can be set up technically, e.g., by barring the access from others than specifically defined set of subscribers.

<u>PLMN on part of MNO's licensed spectrum</u>. Spectrum is typically very expensive investment for the license holder, so dedicating part of it needs to be designed carefully applying techno-economic optimization, in order to adequately balance the cost and expected quality.

<u>SA NPN on unlicensed spectrum</u>. The radio deployment and RAN business case are light as there is no license fee involved. Nevertheless, the QoS cannot be ensured due to load of shared spectrum with possible other users.

### *B. PNI-NPN deployment*

The following list summarizes key aspects of this model.

<u>NPN shared RAN</u>. NPN and PLMN share part of the RAN, but the NPN communications stay within the defined premises. 3GPP defines well the technical RAN sharing options that can be applied in this model.

<u>NPN shared RAN and CP</u>. NPN and PLMN share the RAN for the defined premises while the PLMN has network control; the NPN traffic remains within the defined premises. Network slicing serves this model as per 3GPP specifications, complemented by industry forums' guidelines for slice template setup. Alternatively, the setup can be based on APN.

<u>NPN hosted by public network</u>. PLMN and NPN traffic are external to the business area so that these traffic flows are served by different networks, and the NPN subscribers are effectively public network subscribers.

### *C. Pros and Cons of deployment models*

*a) SNPN*

- Pros: access for customization, independent controlling; high security by full isolation; RAN functions are within reduced geographical area favoring low-latency applications.
- Cons: deployment cost; expertise required for deployment. Dedicated network for sole enterprise includes the cost of the whole system in the geographic area.

*b) Shared RAN*

- Pros: optimizes RAN infrastructure costs while the internal data remains within the NPN infrastructure providing good protection; PLMN RAN connectivity serves for delivering the data meant for outside of the NPN as per need. Within the NPN, part of the base stations can be connected to PLMN

 ISBN: 978-1-68558-068-1 

according to the RAN sharing agreement between the PLMN and NPN operators while the rest can be kept internal. Licensed spectrum copes interferences; deployment less expensive compared to SNPN; uses typically local functions favoring low-latency applications.

- Cons: external interferences can be higher than in SNPN, and the overall control of the network is less independent; need for local expertise, although less than in SNSP.

*c) Shared RAN and CP*

- Pros: licensed spectrum for controlled interferences; lower deployment expenses compared to SNPN and PNI-NPN; SLA can be applied between the NPN and public network.
- Cons: less independent from public networks; latency typically higher than in SNSP and PNI-NPN deployments; some local expertise required.

*d) Hosted solution (Network Slicing by NSP/MNO)*

- Pros: facilitated by NSP, no need for local expertise; fast to set up and adjust based on expressed requirements.
- Cons: less control for adjustments as the NS is managed by the NSP; technology not yet final, practical deployments require SA 5G network that are not yet many in markets.

*e) Open RAN as an NPN service*

- Pros: the cost can be low; easy to set up by provider; can be hosted as "light-weight" 5G SNPN or NPI-NSN.
- Cons: technology is still evolving and the realistic products for Open RAN -based NPN can take time.

*f) 5G FWA (home office use case)*

- Pros: Customer Premises Equipment (CPE) easy to install; replaces fixed cabling.
- Cons: use case adequate in a home office environment, but limited for larger enterprise NPN use.

*g) Non-5G-based solutions (Wi-Fi hotspots)*

- Pros: Wi-Fi hotspot deployment is rather straightforward within an enterprise area. The radio coverage does not require license, and it can also be extended to reach any Wi-Fi device external to the enterprise premises.
- Cons: low security, limited mobility, and lack of QoS.

*h) LPWAN*

- Pros: many options available basing on both cellular and non-cellular radio technologies. Cellular-based LPWANs are services integrated to system, easy to deploy.
- Cons: non-3GPP-based LPWANs have varying security and protection levels, and they require a separate infrastructure.

## VI. TECHNO-ECONOMIC OPTIMIZATION MODELING ASPECTS

### A. Current deployment models

Each NPN deployment model has their advantages and disadvantages. Understanding the requirements of the use cases, and applying tecno-economic optimization assessment that considers the key variables, the selection of the most adequate deployment model will benefit favorable business.

In the selection of the deployment model, the task is to understand the realistic needs of the final users of the NPN access, performance, security aspects, mobility, capacity, QoS, and other key factors, and how they can be served by applying cost-efficient technical solutions. It is important to understand also the changing requirements in the foreseen future because along with the evolution of the environment, originally selected, initially optimal model, can turn out to be less optimal in longer term.

As described in [11], related to NS scenarios, the collection of the verticals' requirements can be done via practical means to interpret the needs of the end-users. From the operators' perspective, this can be somewhat challenging task as the verticals may often express their requirements using non-standard terminology, or are not able to formulate the actual requirements. The methodology in [11], despite its original focus on NS, can be useful to be extended to cover additional aspects the other NPN deployment models. This methodology suits to interpret practical vertical needs based on the foreseen use cases, and to form technical requirements that can be finally be mapped to represent input parameters for the NPN modeling, whether it is about tailored 5G NS template for which the GSMA PRD NG.116 serves as a base, more "traditional" standalone NPN setup, or shared network model. The aim of the requirement list is to ensure the common understanding of the environment, and set the expectations also for service level.

### B. On further techno-economic optimization

The assumption of the modeling is that the NPN is a business between an entity capable of deploying adequate wireless network (such as MNO, NSP, network equipment vendor, or system integrator) and an enterprise desiring to facilitate the mobile communications for their end-users in their communications via Industry IoT applications (such as port or energy company monitoring and controlling their workflow via intelligent sensor network).

The model for the selection of the most adequate NPN deployment option can build upon modular elements, which are:

- *Interpretation* of the enterprise and end-user needs (e.g., via survey) to form technical requirements statement as a base for the input parameters; e.g., capacity (number of expected users/devices), coverage, QoS, need for roaming/local-only utilization; and services (IoT, voice, other).
- *Business aspects* assessment (understanding possibility for investment in terms of CAPEX and OPEX; flexibility for initial and longer-term investments).
- *Forecast* of today's, near-future, and longer-term outlook for the possible need for expansion of the network, capacity, and evolving QoS (which is important to avoid investing to multiple types of NPN as the requirements evolve).
- Any other relevant information on the deployment aspects.

The assessment of the feasibility of the deployment options relevant to the scenario under evaluation can be carried out based on these results in a comparative manner. The base for the economic assessment is the cost for enterprise in terms of the CAPEX (initial deployment and forecasted posterior need for new infrastructure investments) and OPEX (yearly cost in order to operate NPN). The cost estimate of each scenario presented in the NPN Deployment Models Section considers the key attributes, such as area of deployment, device number (total expense $x_d$), and radio performance indicators, which together result in the required bandwidth and number of radio cells, and finally in the total cost of cells $x_{gnb}$. As an example, in very high data rate scenario requiring large indoor and outdoor NPN, the number of mm-Wave small cells, each resulting in, e.g., 80-100 m cell range, can be in order of dozens per km$^2$. Other attributes consider the cost of transmission network $x_{tn}$, and spectrum $x_s$.

 ISBN: 978-1-68558-068-1 

For the core network, the cost includes the licenses and other cost items to activate the needed network functions NF $x_{nf}$. Cost of the needed applications / services $x_a$ refers to the support of, e.g., voice service (that requires either own or outsourced IMS core for integrated Voice over New Radio) and IoT service license, and Location Based Services (LBS) deployment. Roaming and interconnectivity cost $x_r$ is related to the agreements with national and international networks.

The cost $x_v$ of other variable items can include, e.g., the actual installation of RAN, Transmission Network (TN), and Core Network (CN) equipment or cloud environment, including antenna systems, base station shields, cabling, and any other expense that is required to set up the NPN for enterprise.

For the estimate of yearly operating costs $y$, the same main components as presented in CAPEX analysis generate expenses, such as licensing fees and electricity consumption, whereas an additional item to be considered in operations is the maintenance cost $y_m$. The resulting Equations for the initial costs (1) and operating costs (2) are thus

$$CAPEX = x_{gnb} + x_{tn} + x_s + x_a + x_d + x_v + x_r + \sum_{nf1}^{nfn} x_{nf} \quad (1)$$

$$OPEX = y_m + y_{gnb} + y_{tn} + y_s + y_a + y_d + y_v + y_r + \sum_{nf1}^{nfn} y_{nf} \quad (2)$$

It should be noted that the variables in Equations (1) and (2) can have non-linear inter-dependencies; as an example, the volume discount of the number of radio cells can also lower the relative cost of core software and cloud feature licensing.

The assessment results in a statement of the suitability of deployment scenarios to indicate their level of compliance with the requirements. This method can be visualized in terms of the total cost per area as a function of time, considering attributes of interest, such as maximum supported device number or maximum data rate. The method serves thus to estimate the initial and longer-term cost of each deployment model under evaluation; it is possible, that the initially most cost-efficient option might turn out to be less optimal in longer run.

### C. Return on Investment

The described modeling can be extended to estimate Return on Investment (RoI) of private network, including the business of MNO, enterprise, or 3rd party. The RoI depends on the deployment and operational costs, share of ownership of private network components (hardware, software) versus outsourced items (e.g., 5G core that runs in virtualized environment served by cloud provider) in different deployment scenarios of interest, the generated savings compared to reference deployment scenario (as an example, enterprise can compare MNO-operated scenario against completely or partially enterprise-owned network), as well as potential earnings for different stakeholders. As an example, enterprise managing completely or partially owned private network, either on shared or own spectrum, could allow also additional users to roam into that network for a fee that depends on the data consumption or time.

Although the pricing of network components is business between the vendors and customers and thus largely non-public information, it can be assumed that large entities investing to either own network infrastructure or outsourced solutions to provide the private network services to the end-users, may benefit from lower costs due to scale of economies compared to smaller entities. Nevertheless, the private network ecosystem is expected to grow significantly at present, so it can thus also be assumed that small and medium sized entities could reach fortified position for own network component price negotiation in their private network market, which also can impact positively on the pricing models.

### D. Expectations of the modeling

As can be seen from Equations 1 and 2, the selected items result in a linear presentation for initial and operating costs. That said, the Equations represent snapshots of scenarios, and each parameter value may have either linear or non-linear behavior as a function of time, number of components, etc. As an example, the expense related to gNB can be either fixed per the number of gNBs, or there could be a volume-based discount granted by the vendor as a function of the number of gNBs.

As can be expected, the values of the cost items in Equations 1 and 2 depend on the markets, vendor pricing strategies, competitive landscape, and many more variables, so the further analysis of scenarios would be merely speculative without the availability of concrete parameter values. Nevertheless, an example of the potential possible behavior can be presented by testing different scenarios and cost estimates of the parameters to understand the business impact in short, mid, and long-term operation of a private network.

The scenarios can be divided into following categories for the assessment of the total cost of a network, that a) is completely owned, partially owned, or completely outsourced ownership; b) uses licensed, shared, or unlicensed spectrum; c) has no roaming (completely isolated), or has inbound roaming, outbound roaming, or bilateral roaming. To complement the evaluation, additional criteria can be assessed, too, for validating the level of compliance for end-user requirements such as QoS, latency, maximum and average data rate, reliability, etc. The level of compliance of different scenarios can be compared by using numeric values and their weights of importance.

Let us assume an enterprise desires to compare the techno-economic feasibility of a) completely own and isolated small-scale (10 mm-Wave gNBs), 5G network (SNPN) that is based on unlicensed 5 GHz spectrum and 5GC NFs on cloud, with b) MNO-operated private network that is based on an NS dedicated to the enterprise with gNBs that are already partially deployed for PLMN users in the area complemented by new, 5 additional indoor mm-Wave small cells in the enterprise's operational premises. Figure 6 depicts an imaginary example of the relevant key expense behavior over time using the parameters of Equation 1 and 2 and certain estimated values so that they are normalized having the SNPN CAPEX as the reference at year 0.

As can be seen from Figure 6, the initial cost of enterprise's completely own network can be considerably higher than a subscription to an MNO's NS-based service to form a private network due to required investments on the infrastructure. In this scenario, also the OPEX of the SNPN cumulates faster compared to the dedicated MNO NS due to maintenance and licensing expenses of the own network.

 ISBN: 978-1-68558-068-1 

This method provides the enterprise with a tool to estimate the cost difference of deployment models of interest over time and to assess whether certain model is acceptable for deployment regardless of projected, potentially higher cost to balance the key requirements of the enterprise considering, e.g., the level of independent network control and security.

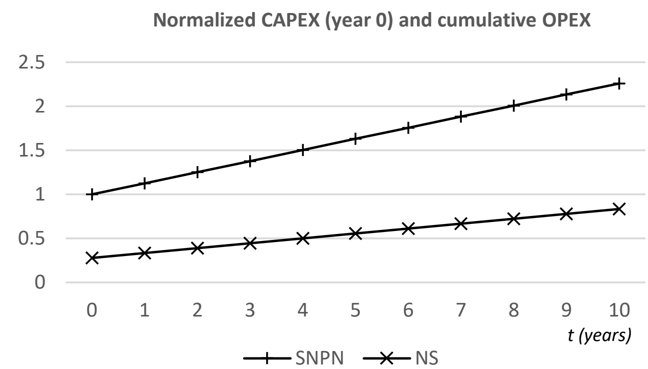


Figure 6 Example of the model's outcome comparing SNPN (reference) and NS deployment scenarios.

## VII. SUMMARY AND FUTURE WORK OPPORTUNITIES

NPN can serve many verticals and their use cases in a more optimal way than PLMN may be able to, to comply with special requirements for, e.g., hardened security by isolation, or high flexibility for network settings adjustment, for which the pros and cons of enterprise-owned vs. operators' components need to be evaluated. There are variety of deployment and ownership models, so the assessment of the scenarios prior to business decisions for the most feasible deployment and ownership model is beneficial.

This paper presents means for the assessment of the techno-economic feasibility of NPN models and an imaginary example on the evaluation. For the model to perform adequately, insights on realistic OPEX and CAPEX values of the model's parameters are important. Thus, feedback from NPN proof of concepts and trials serves to calibrate this modeling and helps identify and focus on the evaluation of the most essential cost items.

The private networks are becoming reality, and they provide a functional base for many verticals and use cases to cope with special requirements. Stakeholders considering deployment and use of private networks benefit from adequate platform. As this study shows, even a relatively simple model can support the ecosystem to better understand the differences of the business cases related to a variety of private network models.

## REFERENCES


[1] GSMA, "Internet of Things," GSMA, [Online]. Available: https://www.gsma.com/iot/manufacturing/private-networks/private-networks/. [Accessed 10 May 2023].

[2] 5GACIA, "5G Non-Public Networks for Industrial Scenarios," 5GACIA, July 2019.

[3] GSMA, "5G industry campus network deployment guideline," 10 November 2020. [Online]. Available: https://www.gsma.com/newsroom/wp-content/uploads//NG.123-v1.0.pdf. [Accessed 10 May 2023].

[4] D. Chandramouli, "5G for Industry 4.0," 13 May 2020. [Online]. Available: https://www.3gpp.org/news-events/3gpp-news/tsn-v-lan. [Accessed 10 May 2023].

[5] H. J. Son, "7 Deployment Scenarios of Private 5G Networks," NetManias, 21 October 2019. [Online]. Available: https://www.netmanias.com/en/?m=view&id=blog&no=14500 . [Accessed 10 May 2023].

[6] 5G-ACIA, "Integration of 5G with Time-Sensitive Networking for Industrial Communications," 5G-ACIA, 2021. [Online]. Available: https://5g-acia.org/whitepapers/integration-of-5g-with-time-sensitive-networking-for-industrial-communications/. [Accessed 10 May 2023].

[7] Syniverse, "Managing the Complexities of Connectivity for Enterprises and Operators," Syniverse, 24 March 2022. [Online]. Available: https://www.syniverse.com/blog/connectivity/managing-the-complexities-of-connectivity-for-enterprises-and-operators/. [Accessed 24 January 2023].

[8] M. Dryjański, "Private 5G Networks – Technology, Architecture, and Deployment," Rimedo Labs, 4 January 2021. [Online]. Available: https://rimedolabs.com/blog/private-5g-networks-technology-architecture-and-deployment/. [Accessed 10 May 2023].

[9] 3GPP, "3GPP TS 23.251 Network Sharing; Architecture and functional description (Release 17)," 3GPP, March 2022.

[10] Samsung, "Technical whitepaper: network slicing," Samsung, 2020.

[11] GSMA, "Network Slicing: North America's Perspective V1.0," GSMA, 3 August 2021. [Online]. Available: https://www.gsma.com/newsroom/wp-content/uploads//NG.130-White-Paper-Network-Slicing-NA-Perspective-1.pdf. [Accessed 10 May 2023].

[12] GSMA, "Generic Network Slice Template V 7.0," GSMA, 17 June 2022. [Online]. Available: https://www.gsma.com/newsroom/wp-content/uploads//NG.116-v7.0.pdf. [Accessed 10 May 2023].

[13] GSMA, "5G Global Launches & Statistics," GSMA, [Online]. Available: https://www.gsma.com/futurenetworks/ip_services/understanding-5g/5g-innovation/. [Accessed 10 May 2023].

[14] GSMA, "Mobile economy report 2022," GSMA, 2022. [Online]. Available: https://www.gsma.com/mobileeconomy/wp-content/uploads/2022/02/280222-The-Mobile-Economy-2022.pdf. [Accessed 10 May 2023].

[15] M. Shelton, "Open RAN Terminology," Parallel Wireless, 20 April 2020. [Online]. Available: https://www.parallelwireless.com/open-ran-terminology-understanding-the-difference-between-open-ran-openran-oran-and-more/. [Accessed 10 May 2023].

[16] O-RAN Alliance, "O-RAN: Towards an Open and Smart RAN," O-RAN Alliance, October 2018.

[17] Intel, "LTE-WLAN Aggregation (LWA): Benefits and Deployment Considerations," Intel, 2016. [Online]. Available: https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/lte-wlan-aggregation-deployment-paper.pdf. [Accessed 10 May 2023].

[18] Ericsson, "Fixed Wireless Access handbook 2023 edition," 2022. [Online]. Available: https://www.ericsson.com/en/fixed-wireless-access. [Accessed 10 May 2023].


 

# Federated Learning for Distributed Sensing-aided Beam Prediction in 5G Networks

Adwitiya Pratap Singh
Hughes Systique Corporation
Gurgaon, Haryana, India
e-mail: Adwitiya.singh@hsc.com

Chitwan Arora
Hughes Systique Corporation
Gurgaon, Haryana, India
e-mail: chitwan.arora@hsc.com

Abheek Saha
Hughes Systique Corporation
Gurgaon, Haryana, India
e-mail: abheek.saha@hsc.com

***Abstract*— The increasing demands for higher data rates have caused newer communication systems to move towards higher frequency bands. However, during the initial network access, the user faces a problem of high beam selection, due to the rich scattering environment and the large number of possible beams. For high mobility and low latency applications, such as vehicular communications, high beam selection overhead is a very big problem. Sensing-aided beam prediction using environmental sensing information as well as telemetry data can be a possible solution to this issue. In this paper, a novel approach is suggested that combines real-time series Global Positioning System (GPS) data, as well as terrain related data for beam selection. Using the DeepSense dataset, we demonstrate that distributed machine learning algorithms, while being computationally tractable, can choose the top N beams with an accuracy that is comparable to that of centralized learning, but faster than it. The novelty of our work lies in the usage of this data set to simulate federated learning and trying different techniques to increase accuracy.**

***Keywords-Wireless Technology; Artificial Intelligence; Deep Learning; Federated learning*.**

## I. INTRODUCTION

Current and future communication systems are moving to higher frequency bands. The large available bandwidth at the high frequency bands enables these systems to satisfy the increasing data rate demands of the emerging applications, such as autonomous driving, edge computing, and mixed reality [1]. These systems require the deployment of large directional antennas at both the Transmitter (Tx) and Receiver (Rx). Using directed beams to connect to the network introduces a new problem, which is choosing the optimal beam from the array of beams present at the transmitter. The overhead for the exhaustive scan to find the beam is way too high for applications that need low latency, hence we have our pain point. The way we are moving towards solving this problem is machine learning for optimization and forecasting the beams.

Sensing aided beam prediction seems to be the foot in the right direction: The mm-wave communication dependence on Line-Of-Sight (LOS) links between Tx and Rx really brings into play the sensory aid that can be provided by sensors on the transmitter and the receiver side. With the aid of GPS and image sensors, the transmitters can decide in which direction to point their beams by seeing the traffic distribution and identifying the receivers through visual sensors. This will narrow down the search done by the exhaustive scan during the initial access (as described in [2]).

Recent work on sensing-aided beam prediction has shown unprecedented results in using the sensory data, such as Red, green, blue (RGB) images, LIDAR, radar, and GPS positions for the beam prediction problem. However, the previous research is mainly done on synthetic datasets (datasets which have data that have been simulated or created virtually). While these datasets provide us insight into how the real time model would perform, there is still a disparity between modelled performance and real-time performance. Some features, such as obstruction and time of day can only be simulated on a real time dataset. This is what we are trying to achieve in this dataset. In this research paper, we will commence by reviewing the previous work conducted in this area in Section 2, followed by an in-depth examination of the problem in Section 3. Section 4 will focus on the discussion of federated learning and the distinct aggregation methods employed. Subsequently, we will present our solution implementations and results in Section 5 and Section 6, respectively. We conclude our work in Section 7.

## II. RELATED WORK

There has been previous work done on synthetic dataset. In [3], the authors found out that in a raytracing implementation the deep neural network model was able to accurately predict the beam parameters up to 90%. The paper also investigated how multiple Remote Radio Heads (RRH) working together could be used to increase prediction accuracy and how they could be implemented using a hybrid edge cloud model.

The authors of [4] analysed multiple types of Deep Neural Networks (DNN). With the use of multi-modal data such as LIDAR and using separate machine learning models, they achieved an accuracy of 91.2%.

The authors of [5] propose a beam selection model based on Convoluted Neural Network (CNN). Their CNN model for the latter should contain 6 layers (2D) and 1 linear layer. GPS data was used at this point, plus the added four linear

 ISBN: 978-1-68558-068-1 

layers. Their results showed 96.9% accuracy in the top 10 accuracy.

The authors in [6] focused on LIDAR data. They also explored the use of federated learning using different clients as well as using CNNs on the different nodes. The LIDAR used was from mounted sensors of the vehicles.

The authors in [7] tested the federated network used for mmWave beam-selection against a backdoor attack algorithm. Their attack basically consisted of creating obstacles on the road at specific locations. The main purpose of the attack was two-fold (1) to force the model to output a beam in a desired direction (2) to send a low signal strength beam.

In [19], the authors propose a distributed learning framework that leverages multiple vehicles as clients, each equipped with mmWave communication capabilities. The paper explores the effectiveness of this approach and demonstrates its ability to achieve accurate beam selection in vehicular mmWave systems.

Compared to the previous research, the novelty of our work is two-fold. As opposed to the works cited above, we have combined time series GPS data and stacked it with image data from the infrastructure and measure the beam selection accuracy. This will provide us with results that we can expect during practical deployment.

## III. PROBLEM OVERVIEW

In an mm-wave wireless network, beamforming is an important technique used to improve the efficiency and capacity of the network. Beamforming involves adjusting the directionality of the antenna beams to focus the signal towards the intended receiver, rather than broadcasting it in all directions. This technique can be particularly effective in dense urban environments, where there are many obstacles and scattering sources.

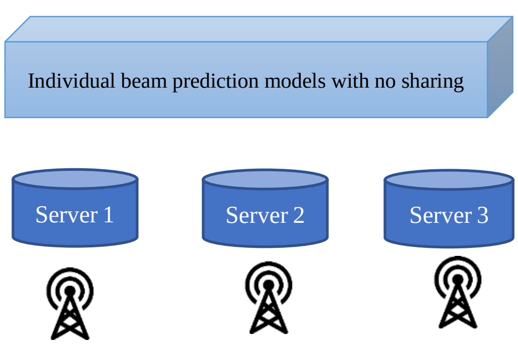


Figure 1. Centralized Beam-selection using AI.

The selection of the optimal beam for a given user is a rich target for machine learning based algorithms since there is no deterministic way to achieve this other than an exhaustive search [8]. In the first generation of machine learning algorithms, as shown in Fig. 1, gNodeB would be running in isolation. The data would be fed separately, and models would not be trained on each other datasets which would not allow the models to be apprised of different traffic distributions that are viewed by neighboring gNodeBs.

In a real-life deployment, each gNodeB would have a view of only a specific location/scenario. To create and run a centralized model, the gNodeB should have access to all possible data or scenarios. However, it would be prohibitively expensive to get the data in one centralized place. Therefore, the practical solution would be to move towards a distributed model. To properly allocate the beam index in a 5G deployment, a gNodeB can use the data of other gNodeBs around it. By sharing information with nearby gNodeBs, each can better understand the overall network conditions and adjust its beamforming accordingly. Data sharing requires nodes to collaborate even though their data may be different in terms of distribution, quality, and quantity.

However, there are significant challenges to the data sharing approach as well. Non-linear aggregation can cause the model to move in the opposite direction of the actual convergence point, so an optimized aggregation technique must be implemented. Sharing data in real-time also requires a certain amount of bandwidth that sometimes cannot be allocated due to external non-controllable factors [9]. Data sharing also brings about the risk of breaches in the network as discussed in [10].

In the context of V2I (Vehicle to Infrastructure) communication, data sharing has several advantages over centralized learning, such as the reduction volume of data and consequently latency. Furthermore, since each device can participate in the training process without the requirement for a robust central server, computing resources are employed more effectively. For instance, gNodeB can use a trained model to choose the appropriate beam for each car in an area where it detects numerous vehicles. Like this, gNodeB can utilize its learned model to modify its beamforming to avoid a certain direction if it detects high interference in that direction.

## IV. FEDERATED LEARNING

### A. Approach to using Federated Learning

Federated learning addresses the challenges as mentioned in the problem overview section. In our solution, as shown in Fig. 2, the gNodeB will identify the nearby towers and let them know they have similar data that can aid their beam prediction algorithms as well. There are known challenges for handling different modalities of data and optimizing the training process for every node is a big challenge [11] which we discuss in the next section.

 ISBN: 978-1-68558-068-1 

*B. Some downsides and their solutions*

Expensive communication is a huge bottleneck in federated learning networks. Since there are millions of end devices that are usually connected in the network that at any time might be aiding the global model, the computation can be slower by many magnitudes [12]. Another problem is system reliability; if the network is comprised of many end devices such as vehicles, at any given time during a training procedure, a local device can dropout due to local system failure [13], which can lead to spurious results. Another issue is scalability, which also plays a big role, as we cannot waste much time during image feature extraction. To mitigate these issues, our task is to develop an efficient extractor that is not computationally intensive. To reduce the computation magnitude, we tried one of two methods: (1) to reduce total communication rounds of federated learning and (2) to try to reduce the data that is being transferred in the network.

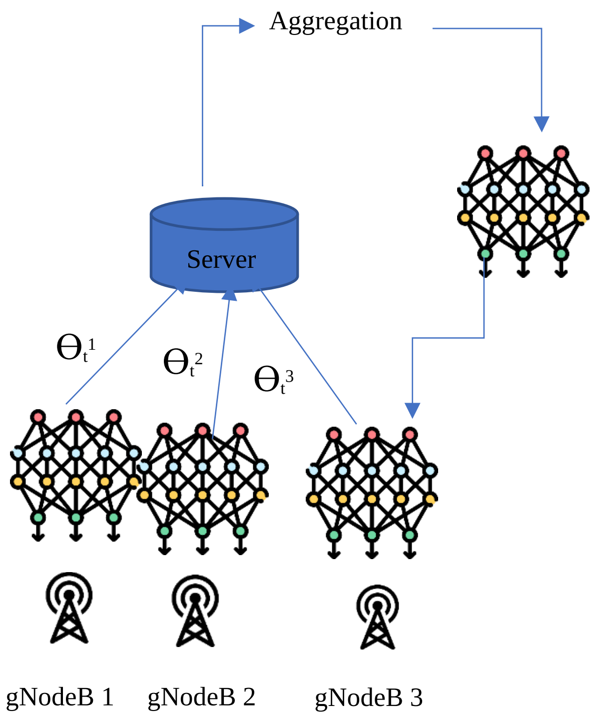


Figure 2. Federated learning from data of multiple gNodeB global model creation using aggregation techniques.

Hence, when we take the gNodeB as the end devices, we are just transferring GPS coordinates in the federated network from the vehicles to the base stations, greatly reducing the size of data transferred than if we kept vehicles as local devices. This will also increase system reliability since the chances of one gNodeB going down is significantly lower than the failure of a vehicle. To increase the scalability, we are using a computationally efficient image extractor rather than heavy transfer learning models to extract features from base station images.

*C. Aggregation techniques used*

While there are many different aggregations models, the best working aggregation model worked with the algorithm used by Federated Stochastic Gradient Descent (FSGD). FSGD is an alternative to averaging in which the client models are updated using Stochastic Gradient Descent (SGD) [14] before sending them to the server for aggregation. The server then combines the updated models using a weighted average. Following are the steps in this aggregation technique:

*1) Local Computation*

Initially each model is initialized with the same weights rather than independent initializations since according to this article [15], common initialization causes better results. The base station acts as the local server where the deep learning takes place using SGD, also the place where the cars share their GPS locations for training.

*2) Model Update*

In this step, each party sends its local model update to a central server. The updates are typically compressed using techniques like quantization to reduce communication overhead. The global model update is given by:

$$\Delta w(k+1) = [1]K * \textstyle\sum_{i=1}^{k} (N_i|N)^* \Delta w_i(k) \qquad (1)$$

Where in equation (1) delta $w_i(k)$ is the local model update of party 'i' at iteration k, N is the total size of the data held by all parties, and K is the number of parties. The weights ($N_i$ / N) ensure that parties with more data contribute more to the global update.

In this step, according to the figure each local server or base station needs to send the local model to the central server, this happens after all the epochs in that round of every client is completed.

*3) Aggregation*

In this step, the central server aggregates the global model update and sends the updated model parameters back to the parties as shown in equation (2). The aggregation can be done using different methods, such as weighted averaging, FSGD, proximal and others that have a higher privacy measure. In the case of FSGD after the weighted average is formed of the given clients then the difference between the current global model weights is computed after which we subtract the difference to move opposite to the rising gradient and towards the convergence.

$$w_{t+1} \Leftarrow \sum_{k=1}^{k} \frac{n_k}{n} w_{t+1}^{k} \qquad (2)$$

We contrasted FSGD against two other techniques described below.

Federated Averaging with Momentum (FedAvgM): This is an extension of the FedAvg technique that includes momentum in the aggregation step. The idea is to maintain a

 ISBN: 978-1-68558-068-1 

running average of the model weights across multiple rounds to improve convergence.

Federated Proximal: Federated Proximal is a technique that uses a proximal operator to enforce sparsity in the model updates. The proximal operator is applied to the global model parameters before they are sent to the clients, and to the client updates before they are sent back to the server. This helps to reduce the communication overhead and improve the efficiency of the federated learning process.

## V. IMPLEMENTATION

### A. Dataset

We have implemented our model on the use case of vehicle-to-infrastructure, specifically scenario (32-34) according to the DeepSense6G dataset [16]. The testbed for getting data for these scenarios has two units: Unit 1 (a stationary unit), which acts as the base station, and Unit 2 (a vehicle), which represents the mobile user. Unit 1 is equipped with the following devices:

1) *A mmWave reciever*
2) *RGB Camera*
3) *3D LIDAR*
4) *Radar*
5) *GPS*

A scenario is a dataset collected from a combination of a transmitter (deepsense testbed 1) and receiver (vehicle) at a certain location. These scenarios differ from each other in terms of either their location or time of day. We use the different scenarios to get the independent behavior of the gNodeB.

Each scenario is a temporally ordered combination of multiple types of data, which is recorded in every 100ms. Corresponding to every timestamp there are 5 instances of image data and 2 instances of GPS data. Our algorithm will exploit the temporal information in the dataset using Gated Recurrent Units (GRUs) will be explained in the further section.

### B. Model

The model receives two inputs: a sequence of position coordinates and an image. After batch normalization and Rectified linear activation unit (ReLU), the image is run through a CNN with four convolutional layers. After being flattened and passing through a fully connected layer with 128 units, batch normalization, and yet another ReLU activation function are applied to the output of the final convolutional layer.

The position coordinates are routed via a GRU layer as shown in Fig. 3 with two layers and 64 hidden units after being first embedded using a linear layer. A fully connected layer with 64 units receives the output of the GRU layer at every time step. Here we use the gated recurrent unit for processing position data since this data is temporally corelated. This step allows us to gauge the movement of the car in play. Long Short Term Memories (LSTMs) were also considered in this step but as our aim was to make this model as computationally inexpensive as possible, we went forward with the GRU, as shown in the figure below as well as the baseline solution [17].

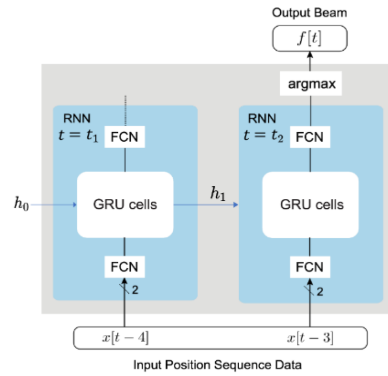


Figure 3. Visual representation of the GRU architecture used to learn the GPS data [9].

The outputs of the CNN and the GPS model are concatenated, and then passed through another fully connected layer with 128 units, followed by another ReLU activation function. Finally, the output is passed through a linear layer with number of classes units, which produces the final classification output.

The model uses dropout regularization with a rate of 0.5 to prevent overfitting, and batch normalization to speed up training and improve the model's ability to generalize to new data.

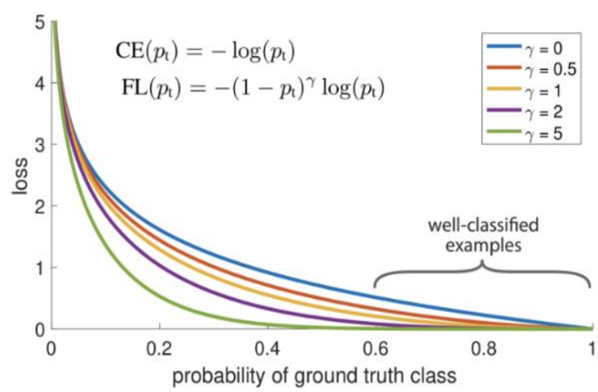


Figure 4. Focal loss representation of changing the modulating factor gamma on the loss [22].

 ISBN: 978-1-68558-068-1 

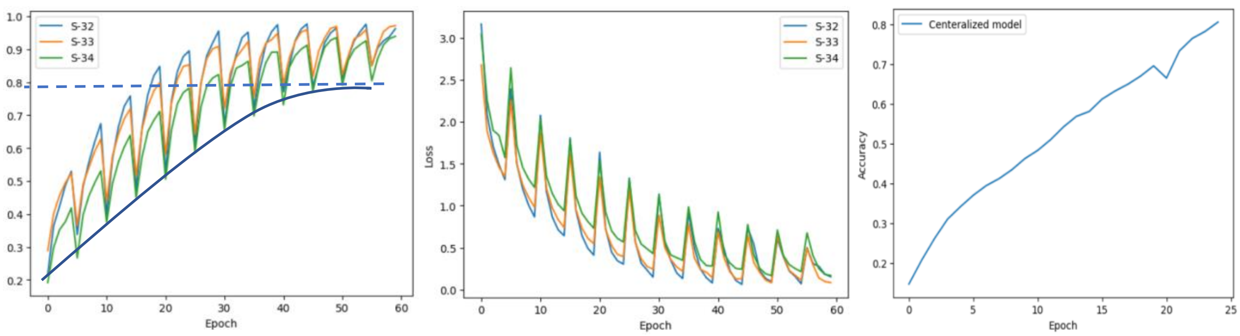


Figure 5. (a) Displays the accuracy chart of the federated learning model through all the rounds (b) shows the decreasing loss of the same federated learning model (c) the accuracy of model that has the same architecture as model before but in centralized environment.

## VI. RESULTS

There are three phases in the solution that predicts the optimal beam index based on the multi modal data. In the first, data collection phase the data from local vehicles to be sent to gNodeB. Over here, the local gNodeB are initialized with the same model, without any fine tuning or aggregation from other gNodeB. In the second model update phase, every gNodeB shall send their model to the central server to be aggregated using federated stochastic gradient descent. All the models shall be aggregated to be sent to the local gNodeB for the next round. We tested out multiple local epoch numbers and came to the optimal number of 5 epoch per round. In the last phase, the updated model is sent to the local gNodeB to help implement the beam selection using the local data.

We can see in Fig. 5 that, after 10 rounds (each containing 5 epochs), the accuracy seems to be stagnating, as noted by the best fit line. We are emphasizing the minimum accuracy amongst the peaks since that is the result after aggregation. This dip in accuracy is due to the new scenario data weights that is introduced to the global model, it maxes out at 80% accuracy in beam selection. To understand why this happened we compared the baseline model to the centralized as well as centralized multi modal model to find the disparity that we will face in accuracy.

The centralized implementation is identical to the federated model except that we used the entire dataset at one node to train the model at once. Since we have a non-IID dataset this is better in terms of accuracy. But as we move towards the real-world application, the processing time consumed in training the entire dataset at once will incur a high latency. As seen in Table I, the best federated model results do lag the centralized model, but it covers in time to process, since parallel processing of three models at three different nodes allowed the model to train 37% faster on the CPU. This would be increased even further if the data is loaded on to the GPU.

A further consideration is that in every scenario there was a different amount of data available to it, and since the data was already non-Independent and identically distributed (non-IID) we used focal loss to penalize our model. In the focal loss as seen below the modulating factor reduces the contribution to the loss from easier examples such as ones which have high frequency in the dataset and extends the range in which an examples receive low loss [18]. We kept the modulating factor to 5 (shown in Fig. 4) as it provided us with the best results.

TABLE I. RESULTS OF DIFFERENT MODELS USED.

| ***Models*** | ***Top 5 of 64*** | ***Top 10 of 64*** |
|---|---|---|
| Baseline model (GRU) | 77 % | 80% |
| Centeralized model | 83% | 90% |
| Federated Model | 64% | 80% |

TABLE II. RESULTS OF DIFFERENT AGGREGATION TECHNIQUES.

| ***Models*** | ***Top 5 of 64*** | ***Top 10 of 64*** |
|---|---|---|
| FSGD | 64 % | 80 % |
| Proximal | 60 % | 75 % |
| Fed Avg | 65 % | 76 % |

In Table II, we can compare the different aggregation techniques used during federated learning. As mentioned before in the implementation section, we know that the federated stochastic gradient descent worked best amongst all. This can be corroborated with theory as well since FSGD is slightly immune to the non-IID imbalanced dataset since it allows for more local model updates. The use of sampling only a subset of the local data to perform the local updates helps FSGD pay less attention to outlier data, as well as making the gradient correct. This is very important when not using such a large dataset such as ours, as well as having a small number of nodes. Although the performance could be further improved if we were able to introduce more types of scenarios of V2I from the Deepsense dataset hence increasing our number of nodes.

## VII. CONCLUSION AND FUTURE WORK

In this paper, we have demonstrated the use of FSGD based federated learning optimal beam selection. We used aid from sensors to allow a multi-modal model accurately predict the beams. The use of other sensors can also prove to be a viable option in sensor-aided beam prediction such as accelerometers and gyroscopes. Different aggregations techniques can be explored to analyze the resultant effect in the performance of federated learning. The federated model falls prey to overfitting if given a small number of clients, we can investigate the behavior by varying the number of active clients in federated learning.

In conclusion, sensing-aided beam prediction is a promising solution for the challenges faced by mmWave communication systems. The utilization of sensory data collected by various sensors can guide the beam management process and significantly reduce beam training overhead. In real-life deployment it is impractical to get all the data at one centralized place for training as a result federated learning can be used as a preferable training solution. Although it is noticed that the accuracy of the federated model is lesser than that of centralized model, we can see that we have a trade of between accuracy and practical realization of latency. Our work received a top 10 running accuracy score of 80%. Federated stochastic gradient descent produced the best results in terms of aggregation techniques.

## REFERENCES

[1] ITU AI for Good Global Summit ", AI for Good Global Summit 2021", ITU AI For Good Challenge, https://challenge.aiforgood.itu.int/match/matchitem/72

[2] H. Lindgren and M. Hasselqvist, "The Effects of Battery Aging on the Energy Management of a Parallel Hybrid Electric Vehicle", Proceedings of the 2017 19th European Conference on Power Electronics and Applications (EPE'17 ECCE Europe), Chalmers University of Technology, 2017, pp. 1-6.

[3] C. Arora and A. Saha, "AI Based Beam Management for 5G (mmWave) at Wireless Edge", IARIA International Journal on Communications (IJCN), vol. 11, no. 3 & 4, 2021, pp. 44-53. DOI: 10.1016/j.ijcon.2021.100038

[4] J. Smith et al., "A Novel Approach to Federated Learning", IEEE Signal Processing for Wireless Communications (SPAWC), IEEE, 2019, pp. 1-6. DOI: https://doi.org/10.1109/SPAWC.2019.8815569Sdad

[5] J. Yang et al., "Federated Learning for Intelligent Vehicular Networks: Challenges and Solutions", IEEE Transactions on Vehicular Technology, vol. 71, no. 3, 2022, pp. 3079-3091. DOI: https://doi.org/10.1109/TVT.2022.3142513

[6] S. Ashish et al., "Secure Federated Learning for Next-Generation Wireless Communication Networks", IEEE 17th Annual Consumer Communications & Networking Conference (CCNC), 2021, pp. 1-6

[7] M. Akbari and E. Mohammadi, "Risk and Advantages of Federated Learning for Health Care Data Collaboration", ResearchGate, 2020.

[8] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Gated feedback recurrent neural networks", arXiv preprint arXiv:1502.02367, 2015.

[9] Y. Chen, J. Liu, M. Xu, and K. Ren, "Federated learning in mobile edge networks: A comprehensive survey", IEEE Communications Surveys & Tutorials, vol. 20, no. 3, 2018, pp. 2022-2063. DOI: 10.1109/COMST.2018.2834990

[10] M. Karimzadeh et al., "Secure Federated Learning with Homomorphic Encryption: A Review", IEEE Access, vol. 9, 2021, pp. 197789-197814. DOI: https://ieeexplore.ieee.org/document/9862973

[11] "Deep Learning-Based mmWave Beam Selection for 5G NR/6G With Sub-6 GHz Channel Information: Algorithms and Prototype Validation", ResearchGate, 2020, DOI: 10.13140/RG.2.2.15928.44809

[12] C. Van Berkel, "Multi-core for mobile phones", In Conference on Design, Automation and Test in Europe, 2009, pp. 1-6.

[13] A. Kumar Sahu, T. Li, A. Talwalkar, and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions", Carnegie Mellon University, Bosch Center for Artificial Intelligence, and Determined AI, 2019.

[14] J. Konečnỳ et al., "Federated learning: Strategies for improving communication efficiency", arXiv preprint arXiv:1610.05492, 2016.

[15] "CS294-163: Federated Learning", University of California, Berkeley, Electrical Engineering and Computer Science, https://inst.eecs.berkeley.edu/~cs294-163/fa19/slides/federated-learning.pdf.

[16] G. Charan Alkhateeb et al., "DeepSense 6G: large-scale real-world multi-modal sensing and communication datasets", to be available on arXiv, 2022.

[17] DeepSense6G, "Multi-Modal-Beam-Prediction-Challenge-2022-Baseline", [GitHub repository], https://github.com/DeepSense6G/Multi-Modal-Beam-Prediction-Challenge-2022-Baseline, 2022.

[18] T. Y. Lin et al., "Focal loss for dense object detection", In Proceedings of the IEEE international conference on computer vision, 2017, pp. 2980-2988.

[19] M. Brown, C. Thompson, and B. Anderson, "A federated learning approach for beam selection in vehicular mmWave systems", In Proceedings of the IEEE Global Communications Conference (GLOBECOM), 2021, pp. 1-6.

 ISBN: 978-1-68558-068-1

# Movement Generators in Mobile Medium Ad Hoc Network Simulation

Przemyslaw Pochec
Faculty of Computer Science
University of New Brunswick
Fredericton, Canada
e-mail: pochec@unb.ca

***Abstract*—Mobile Medium Ad Hoc Network (M2ANET) is the network model introduced in 2011 that could replace the Mobile Ad Hoc Network (MANET) model. Rather than focusing on user node mobility, the M2ANET models a cloud of mobile nodes forming a forwarding network accessible to any users (stationary or otherwise). The performance of such a network depends on the pattern of node movements. Thus, any performance evaluation requires modelling of the movement of the forwarding nodes. We review different node movement paradigms used in our research and suggest use cases for their applications.**

***Keywords-MANET; Mobile Medium; simulation; movement generators.***

## I. INTRODUCTION

The Mobile Medium Ad Hoc Network (M2ANET) concept [1] introduced in 2011 precedes the discussion of the drone networks [2] and Flying Ad Hoc Networks (FANETs) [3]. As a new concept, it introduces a cloud of mobile forwarding nodes at a service of communicating clients. The Mobile Medium is not necessarily tied to any physical implementation of mobility so, as a model, it can be applied in multiple scenarios, including the likes of drone and FANET networks. Simulation of M2ANETs requires modelling of the movement of mobile nodes [4]. In this short paper, we survey the node movement generation techniques used in our research on M2ANETs over the past decade.

In Section II, we review the new Mobile Medium model and compare it to a standard MANET. In Section III, we present different methods for modelling movement of mobile nodes in simulation. In Section IV, we discuss a novel methodology for node movement modelling which is based on processing the movement files themselves. Finally, Section V presents some ideas about the future of Mobile Medium research.

## II. MANET VS M2ANET

Typically, a network is called a Mobile Ad Hoc Network if it consists of a group of mobile wireless nodes exchanging messages with one another [5]. During the lifetime of the network, nodes move freely and form opportunistic connections that allow for establishing routes and forwarding data within the network. Thus, the node mobility plays an important role in formation of the links, establishing the routes and then forwarding data between the nodes [5]. In a Mobile Medium Ad Hoc Network [1], the mobile nodes are divided into two categories: (i) the forwarding only nodes forming the so-called Mobile Medium, and (ii) the communicating nodes, mobile or otherwise, that send data and use this Mobile Medium for communication. The key performance measure of the M2ANET is its capability of forming a route between the specific user client nodes sending data. In other words, whether all the mobile nodes forming the Mobile Medium are fully connected is irrelevant as long as they can form a route between the specific users trying to connect through this network. The network is the medium and it only matters for forming a route between the user client nodes.

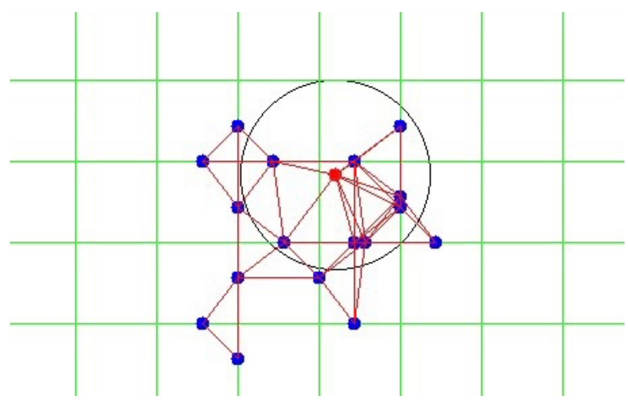

Figure 1. Mobile nodes moving randomly on a lattice.

In a M2ANET, standard MANET routing protocols can be used for establishing the routes for transferring data between the user client nodes through the Mobile Medium created with forwarding nodes [5].

## III. MODELLING NODE MOVEMENT IN MOBILE MEDIUM SIMULATION

In any Ad Hoc Network, opportunistic connections form between stations positioned within the transmission range. In the Mobile Medium, the wireless links proliferate when a sufficient node density is achieved in a region long enough for the routing algorithms to successfully detect the available connections and establish the available routes [1]. The location of the nodes in a network is determined by their movement pattern. Therefore, in order to model the

 ISBN: 978-1-68558-068-1 

operation of the Mobile Medium, we have to model the movement of the nodes. The following is the discussion of the mobile movement models we used in our research.

### A. Random Models

Random mobility is commonly used as a reference scenario in investigating the behavior of mobile networks. It is available in the popular open source simulator ns2 where it is referred to as the Random Way Point (RWP) model [6]. In RWP, nodes are moved in a piecewise linear fashion, with each linear segment pointing to a randomly selected destination (way point) and the node moving at a constant, but randomly selected speed. While the RWP model available in the standard ns2 simulation suite operates in two dimensions, it can also be extended to 3D [7]. RWP models suffer from what is called the border effect [8], which is a non-uniformity in node density occurring along the edges of the region where the mobile nodes are confined to stay. In our research, we experimented with modifications to the RWP movement generator in order to minimize the border effect [9].

### B. Constrained Random Models

In a more realistic scenario, nodes may not have complete freedom to move in any direction. The existing restrictions on the node movement may be due to the physical environment constraints (e.g., obstacles) or due to the limitations on the node propulsion system (e.g., a balloon can change altitude, but cannot move in a chosen direction). An example of a constrained movement modelling would be having nodes placed randomly, but allowed to move only by changing one coordinate, i.e., move either in a horizontal or vertical direction [10]. In our research, we also used another example of the constrained random model with all the node movements constrained to a square lattice, Figure 1, representing a city grid [11]. In this scenario, the nodes can choose to move up and down the lattice grid and can also turn at the grid intersections. We note here that, in similar studies, the network nodes were modelled to move like vehicles in a simulated urban environment, for example in Simulation of Urban Mobility (SUMO) [12].

### C. Deterministic Models

Here, we consider the scenarios where a node is moving in a specific direction predetermined by other factors. This would include the use of the prerecorded movement traces in setting the motion of mobile nodes. One *quasi* deterministic model considered in our research (with possible applications for modelling drone swarms) is based on modelling the movement of nodes in formations [13]. A formation is a group of nodes moving together with one specific node playing the role of a leader; all other nodes in the group follow the leader. In this case, only the leader retains the flexibility of choosing the direction of the movement (e.g., moving randomly) while the other nodes in the group simply retrace the path of the leader node following it at a predetermined distance. Possible parametrizations of this scenario include varying the number of members of the group, the path each member takes and the distance at which one node follows another.

## IV. Processing the Movement Files

In general, network simulations systems use a node mobility model to create network scenarios for testing. First, the movement model is chosen, then the scenario is defined in terms of the number of nodes, the area of movement, initial positions of the nodes, the velocity of their movement etc. Parameters are entered into the movement generator, like an RWP, and the movement file is generated. This movement file is then used as the input to a network simulator, like ns2, to define the networking scenario. The movement file can also be analyzed for characteristics of the movement patterns like, for example, the presence of the border effect already mentioned in the previous section.

In our research on M2ANETs, we proposed a novel approach to generating new movement files by *processing* the existing movement files. (A good analogy to this idea is image processing, where the objective is not to create a new image, but rather to “improve” on an existing one.) This processing can be used to achieve the desired characteristics of the node movements not available directly in the existing movement generators. For example, new way points can be added to the ones already included in the movement file generated based on the RWP model, or the speed of movements can be modified. Specifically, we processed standard RWP generated files replacing each move along a straight line with the movement along a curve [14] (for efficiency, each straight line movement was replaced by the movement along a fractal curve by inserting additional way points into the movement file; the computational complexity of this process was managed by limiting the number of iterations in constructing the fractal curve). A possible application of this technique could be in replacing the straight line trajectories generated using a standard RWP model with a more realistic movement, for example, along ballistic curves in a 3D simulation.

## V. Conclusion and Future Work

Mobile Medium is an interesting way to model the operation of wireless Ad Hoc Networks. The model favors path forming over the full connectivity when assessing the performance of a network. Like in any Ad Hoc Network, the performance is dependent on proximity between nodes which facilitates connectivity and formation of routes across the network. Experimenting with M2ANETs is facilitated by the use of simulators coupled with movement generators.

The standard random movement models allow for establishing general characteristics of the networks built on the principle of the Mobile Medium, like establishing a relation between the node density and the path forming capacity of the Medium. For more particular scenarios, more tailored movement generators may be required.

In this paper, the novel concept of transforming the movement files is introduced. In general, this procedure is applicable to both randomly generated and deterministic (recorded) movement files. Processing the movement files directly would allow to test the new "what if" scenarios without modifying the original process of generating the movement file.

## REFERENCES


[1] J. DeDourek and P. Pochec, "M2ANET: a Mobile Medium Ad Hoc Network", Wireless Sensor Networks: Theory and Practice, WSN 2011, Paris, France, Feb. 2011, pp. 1-4.

[2] E. Yanmaz, S. Yahyanejad, B. Rinner, H. Hellwagner, and C. Bettstetter, "Drone networks: Communications, coordination, and sensing", Ad Hoc Networks, Elsevier, vol. 68, 2018, pp. 1-15.

[3] I. Bekmezci, O. K. Sahingoz, and S. Temel, "Flying Ad-Hoc Networks (FANETs): A survey", Ad Hoc Networks, Elsevier, vol. 11, 2013, pp. 1254-1270.

[4] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for Ad Hoc Network research", Wireless Communications and Mobile Computing, Wiley, vol. 2, 2002, pp. 483-502.

[5] S. K. Sarkar, T. G. Basavaraju, and C. Puttamadappa, Ad Hoc Mobile Wireless Networks, Principles, Protocols, and Applications, Auerbach Publications Taylor & Francis Group, 2007.

[6] H. Ekram and T. Issariyakul, Introduction to Network Simulator ns2, Springer, 2009.

[7] N. Mahmood, J. DeDourek, and P. Pochec, "M2ANET simulation in 3D in ns2", The Sixth International Conference on Advances in System Simulation SIMUL 2014, Nice, France, Oct. 12-16, 2014, pp. 24-28.

[8] C. Bettstetter, H. Hartenstein, and X. Perez-Costa, "Stochastic Properties of the Random Waypoint Mobility Model", Wireless Networks, vol. 10, 2004, pp. 555-567.

[9] R. Alghamdi, J. DeDourek, and P. Pochec, "Avoiding Border Effect in Mobile Network Simulation", The Twelfth International Conference on Networks ICN 2013, Seville, Spain, Jan. 27- Feb. 1, 2013, pp. 184-189.

[10] M. Alzaylaee, J. DeDourek, and P. Pochec, "Linear Node Movement Patterns in MANETS", The Ninth International Conference on Wireless and Mobile Communications ICWMC 2013, Nice, France, July 21-26, 2013, pp. 162-166.

[11] X. Zhang, Simulation of Manet in Urban Environment, University of New Brunswick MCS Thesis, 2010.

[12] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent Development and Applications of SUMO – Simulation of Urban Mobility", International Journal of Advances in Systems and Measurements, vol. 6, no. 3&4, 2012, pp. 128-138.

[13] A. Alshehri, J. DeDourek, and P. Pochec, "The Advantage of Moving Nodes in Formations in MANETs and M2ANETs", The Ninth International Conference on Wireless and Mobile Communications ICWMC 2013, Nice, France, July 21-26, 2013, pp. 228-232.

[14] H. Alseef, J. DeDourek, and P. Pochec, "A Method for Transforming Movement Paths in Wireless Mobile Network Simulation", the International Journal on Advances in Networks and Services, vol. 9, no. 1&2, 2016, pp. 11-19.