6G SNS IA

The Voice of European Industry and Research for Next Generation Networks and Services

6G Security and Trust:

Insights from European SNS JU Projects

Security Working Group

Table of Contents

Table of Figures

Abbreviations Table

Acronym	Explanation
3GPP	3rd Generation Partnership Project
6GCLOUD	Proyecto SNS JU (Smart Networks and Services Joint Undertaking)
6G-IA	6G Industry Association
AOUI	Application Onboarding UI
ApW	Assurance-per-Watt
AS	Autonomous System
CAS	Compliance Assessment System
CNF	Containerized Network Function
CSP	Communication Service Provider
CTI	Confidential Threat Intelligence
DP	Differential Privacy
ELASTIC	Proyecto SNS JU (Smart Networks and Services Joint Undertaking)
eBPF	Extended Berkeley Packet Filter
ENI	Experiential Network Intelligence
ENISA	European Union Agency for Cybersecurity
ETSI	European Telecommunications Standards Institute
FaaS	Function-as-a-Service
FL	Federated Learning
GDPR	General Data Protection Regulation
HAL	Hardware Abstraction Layer
HORSE	Proyecto SNS JU (Smart Networks and Services Joint Undertaking)
ICP	Infrastructure Communication Provider
ICS	Industrial Control Systems
IdP	Identity Provider
IEEE	Institute of Electrical and Electronics Engineers
iTrust6G	Proyecto SNS JU (Smart Networks and Services Joint Undertaking)
ISO/IEC	International Organization for Standardization/International Electrotechnical Commission
ITU-T	International Telecommunication Union Telecommunication Standardization Sector
JIT	Just-In-Time
KPI	Key Performance Indicator
KVI	Key Value Indicator

LIME	Local Interpretable Model-agnostic Explanations
LoTw	Level of Trustworthiness
MANO	Management and Orchestration
MEC	Mobile Edge Computing
MORL	Multi-Objective Deep Reinforcement Learning
MTD	Moving Target Defense
NATWORK	Proyecto SNS JU (Smart Networks and Services Joint Undertaking)
NDT	Network Digital Twin
NFV	Network Function Virtualization
NIST	National Institute of Standards and Technology
PBAC	Policy-Based Access Controls
PDP	Policy Decision Point
PEP	Policy Enforcement Point
PET	Privacy Enhancing Technology
PUI	Human Controllable Privacy UI
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RIGOUROUS	Proyecto SNS JU (Smart Networks and Services Joint Undertaking)
RL	Reinforcement Learning
ROBUST-6G	Proyecto SNS JU (Smart Networks and Services Joint Undertaking)
SAFE-6G	Proyecto SNS JU (Smart Networks and Services Joint Undertaking)
SCO	Security Cost Overhead
SEE	Secure Execution Environment
SHAP	SHapley Additive exPlanations
SMC	Secure Multiparty Computation
SNS JU	Smart Networks and Services Joint Undertaking
SOAR	Security Orchestration, Automation, and Response
SOU	Service Composition UI
STDT	Security Testing Digital Twin
SUNSET-6G	Proyecto SNS JU (Smart Networks and Services Joint Undertaking)
TEE	Trusted Execution Environment
TER	Transparent Evidence Repository
TID	Telefónica Investigación y Desarrollo
TRA	Threat Risk Assessor

UE	User Equipment
UEBA	User and Entity Behavior Analytics
UMU	Universidad de Murcia
VIM	Virtualized Infrastructure Manager
VNF	Virtualized Network Function
VNO	Virtual Network Operator
W3C	World Wide Web Consortium
Wasm	WebAssembly
XAI	Explainable AI
XDP	Express Data Path
ZSM	Zero-touch Service Management

Executive Summary

The shift from vision to architecture for 6G networks introduces profound security and trust challenges that surpass those of 5G, driven by extreme decentralization, AI-native operation, and massive device heterogeneity. Future networks will span a continuum from cloud to edge to device, demanding a move from static, siloed security to cross-layered, continuously monitored, and verifiable mechanisms. This white paper synthesizes the collective work of key European Smart Networks and Services Joint Undertaking (SNS JU) projects—including HORSE, RIGOUROUS, ELASTIC, ROBUST-6G, iTrust6G, 6GCLOUD, NATWORK, and SUNSET-6G—showcasing a critical evolution toward proactive, adaptable, and sustainable security architectures.

A core innovation is the integration of predictive security and assurance into orchestration. Projects like HORSE and 6GCLOUD utilize Security Testing Digital Twins (STDTs) and Network Digital Twins (NDTs) to create sandboxed replicas of operational infrastructure, enabling the validation and optimization of security decisions against threat scenarios *before* deployment. This forms the basis for model-driven, closed-loop security management. Furthermore, projects like ROBUST-6G and iTrust6G redefine trust by implementing autonomous and explainable decision-making. iTrust6G's framework relies on assigning Dynamic Trust Levels based on behavioral attributes and contextual information, proactively denying access to entities exhibiting suspicious deviations, even with valid credentials. These access decisions are transparently logged in a Transparent Evidence Repository (TER), with justifications generated by Explainable AI (XAI), ensuring accountability and auditable governance.

Confidentiality and privacy are addressed architecturally rather than just through policies. RIGOUROUS ensures privacy is a measurable and enforceable attribute by requiring services to use Privacy Manifests during onboarding, and by using SOAR (Security Orchestration, Automation, and Response) loops for continuous runtime monitoring and automatic violation response. For securing processing, ELASTIC and NATWORK champion WebAssembly (Wasm) as a lightweight, sandboxed alternative to containers for deployment across the continuum, especially on resource-constrained edge devices. This Wasm-based execution is coupled with attestation and, where available, Trusted Execution Environments (TEEs) to ensure data and workload integrity during processing. Simultaneously, NATWORK and ROBUST-6G enable confidential collaborative intelligence by leveraging Federated Learning (FL), often reinforced with Differential Privacy (DP) and Secure Multiparty Computation (SMC), to train AI models using distributed data without requiring the centralization of raw, sensitive information.

Finally, the work highlights the emerging need for sustainable security and measurable trustworthiness. SUNSET-6G introduces pioneering metrics such as Assurance-per-Watt (ApW) and Security Cost Overhead (SCO) to ensure security solutions are both robust and ecologically responsible. This concern for sustainability is linked to the development of rigorous metrics, with the SAFE-6G project proposing the Level of Trustworthiness (LoTw) as an overarching Key Value Indicator (KVI) that aggregates

performance across five dimensions: Safety, Security, Privacy, Resilience, and Reliability. These advancements, along with coordinated contributions to standardization bodies like ETSI, 3GPP, and NIST, ensure that 6G security will be integrated, flexible, and capable of verifiable assurance across highly dynamic, multi-domain environments.

1. New Challenges in 6G Security

As 6G moves from vision to architecture, its defining features — extreme decentralization, AI-native operation, massive device heterogeneity, ultra-low latency and sustainable-driven design — introduce profound security and trust challenges beyond those faced in 5G. Unlike previous generations, 6G is not only expected to connect more users and services but also to do so across a continuum that spans from cloud to edge to device, including resource-constrained endpoints, mission-critical systems, and autonomous agents.

Main security challenges to face on 6G must now be embedded in a world where:

- **Networks are dynamic and slice-based**, requiring continuous re-evaluation of trust and security during orchestration.
- **Computation is distributed**, meaning sensitive operations happen outside traditional secure domains.
- **Computation spans resource-constrained endpoints,** introducing complexity in deploying, managing, and securing services across massive device heterogeneity.
- **Services are AI-driven**, raising concerns about the robustness, fairness, and explainability of decisions made by autonomous agents.
- **Entity behavior**, guided by UEBA (User and Entity Behavior Analytics) principles, is contextual, dynamic, and multi-layered, making static access control and fixed policies obsolete.

Emerging projects under the SNS programme — including **HORSE [1]**, **RIGOUROUS [2]**, **ELASTIC [3]**, **ROBUST-6G [4]**, **iTrust6G [5]**, **6GCLOUD [6]**, **NATWORK [7]** and **SUNSET-6G [8]** showcase a shift toward proactive, adaptable, privacy-aware, and sustainable security architectures.

Key innovations developed within these projects are detailed in the following core sections and include:

- Trust and Security Management (e.g., Zero-touch automation, MTD, Explainable AI),
- Privacy awareness (e.g., Privacy onboarding, Federated Learning),
- Confidential computing (e.g., Secure enclaves, Wasm at the edge), and
- Sustainable security (e.g., ApW metrics).

The dynamic, distributed, and resource-sensitive conditions of future 6G networks challenge traditional security thinking, demanding an evolution from static, siloed approaches to cross-layered, continuously monitored, and verifiable mechanisms.

These advancements are substantiated through simulation and testing within the Security Testing Digital Twins (STDTs) and are validated by new KPIs/KVIs for trustworthiness that rigorously measure performance against established thresholds, providing verifiable assurance even before public deployment.

This document summarizes the current landscape and future directions for 6G security based on contributions from these flagship projects, structured around six core areas:

- Trust and Security Management
- Privacy awareness
- Confidential computing
- Sustainable security
- KPI/KVI for security
- Standardization

Each area reflects not only technological advancements but also the cultural and methodological shift toward embedding trust, transparency, and resilience into the design and operation of 6G systems. Table 1 below presents an overview of the primary contributions from the SNS JU projects (HORSE, RIGOUROUS, ELASTIC, ROBUST-6G, iTrust6G, 6GCLOUD, NATWORK, and SUNSET-6G) across the four core security dimensions: Trust & Security Management, Privacy Awareness, Confidential Computing, and Sustainability.

Table 1: Key Contributions of SNS JU Projects to 6G Security

Project	Trust & Security Management	Privacy Awareness	Confidential Computing	Sustainability
HORSE	Secure-by-Design Orchestration via Security Testing Digital Twins (STDT).			
ROBUST-6G	Zero-touch Security Management (ZSM); Explainable AI (XAI) for Trust; Physical-Layer Security.	Federated Learning (FL) with Privacy Budget.	XAI within Secure Enclaves (TEEs) for verifiable inference.	
iTrust6G	Adaptive Access Control based on Dynamic Trust Levels from Behavioral Analysis; Transparent			

	Evidence Repository (TER).			
NATWORK	AI-powered Moving Target Defense (MTD) and Cost-aware Orchestration; Cross-domain Trust Coordination.	FL/SMC for Confidential Collaborative Intelligence; Privacy-aware Policy Negotiation.	Wasm with D-MUTRA (Blockchain Attestation) for integrity verification.	
6GCLOUD	Network Digital Twin (NDT) for predictive security analysis and closed-loop validation.			
RIGOUROUS	Digital Twin Integration; Zero-Trust Orchestration.	Privacy Onboarding via Privacy Manifests; Adaptive enforcement using SOAR loops.		
ELASTIC			Wasm/FaaS architecture with Attestation and eBPF for confidential edge service orchestration.	
SUNSET-6G				Metrics for Security Cost Overhead (SCO) and Assurance-per-Watt (ApW).

2. 6G Trust and Security Management

Five projects contribute to fundamental advances in this domain, where trust and security management must evolve to support fully automated, dynamic, and multi-domain network environments:

The **HORSE** project addresses key security and trust challenges expected in the orchestration of 6G networks, where highly dynamic, automated, and cross-domain service deployment is the norm. In this context, traditional security approaches based on static configurations or siloed validations are no longer sufficient. HORSE proposes a novel secure-by-design orchestration architecture that tightly integrates compliance, risk awareness, and proactive validation across the entire service lifecycle.

At the heart of HORSE's approach is the notion that security and compliance must be embedded directly into the orchestration process. To achieve this, the project develops a Compliance Assessment System (CAS) that evaluates service graphs and deployment plans against regulatory, security, and organizational constraints before any real-world action is taken. This pre-deployment validation is made possible through the use of Security Testing Digital Twins (STDT), which create sandboxed replicas of the operational infrastructure. These digital twins allow orchestration decisions and service behaviors to be simulated, monitored, and tested under a variety of threat scenarios, providing an unprecedented level of assurance and enhancing the system's ability to maintain robust operation against anticipated attacks

Beyond pre-deployment simulation, HORSE enhances orchestration with context-awareness and continuous adaptation. Services are described with machine-readable metadata that includes their security requirements and compliance obligations. This metadata enables the orchestration system to dynamically select and configure network functions that meet the current trust and regulatory conditions. Additionally, real-time telemetry is analyzed to detect deviations from expected behavior or policy, triggering automated re-evaluation of the service graph using the digital twin environment. This continuous, bi-directional telemetry stream is key to keeping the STDT synchronized with the production infrastructure, ensuring the twin accurately reflects the current network state, including resource constraints and configuration of cross-domain infrastructure. The overall process forms a closed-loop orchestration system where trust is continuously assessed and enforced.

HORSE envisions multi-party consortia utilizing this approach, where the STDT can model and validate the security compliance of service segments deployed on shared or collectively owned infrastructure, ensuring unified risk awareness without requiring full visibility into all underlying third-party operational data.

To support this architecture, HORSE introduces several enablers, including dynamic orchestration policies, automated risk assessment modules, and simulation-based verification environments. These components ensure that orchestration decisions are not only efficient and scalable, but also secure and explainable. The orchestration

engine can proactively adjust configurations and re-deploy services based on evolving risk levels, making it suitable for the highly heterogeneous and fast-changing environments expected in 6G.

ROBUST-6G reimagines trust and security management in 6G as an autonomous, explainable, and distributed process. Recognizing the limitations of static or manually enforced security policies, the project proposes a macroservice-based architecture that enables security to be orchestrated and adapted in real time, as part of a fully automated service lifecycle. At the heart of this approach is the integration of Zero-touch Security Management (ZSM), Federated Learning, and Explainable AI (XAI), offering a future-ready alternative to traditional security operations.

As shown in Figure 1, the architecture proposed by ROBUST-6G is based on macroservices acting as autonomous agents distributed across the continuum (edge, cloud, and devices). These agents sense and analyze their local environments using Federated Learning, allowing the agents to collaboratively train and update aggregated security models. They then execute threat mitigation actions in real time and explain their decisions through integrated Explainable AI modules. This enables a zero-touch security platform where orchestrated security decisions are both autonomous and auditable, while preserving privacy through decentralized processing and physical-layer trust signals.

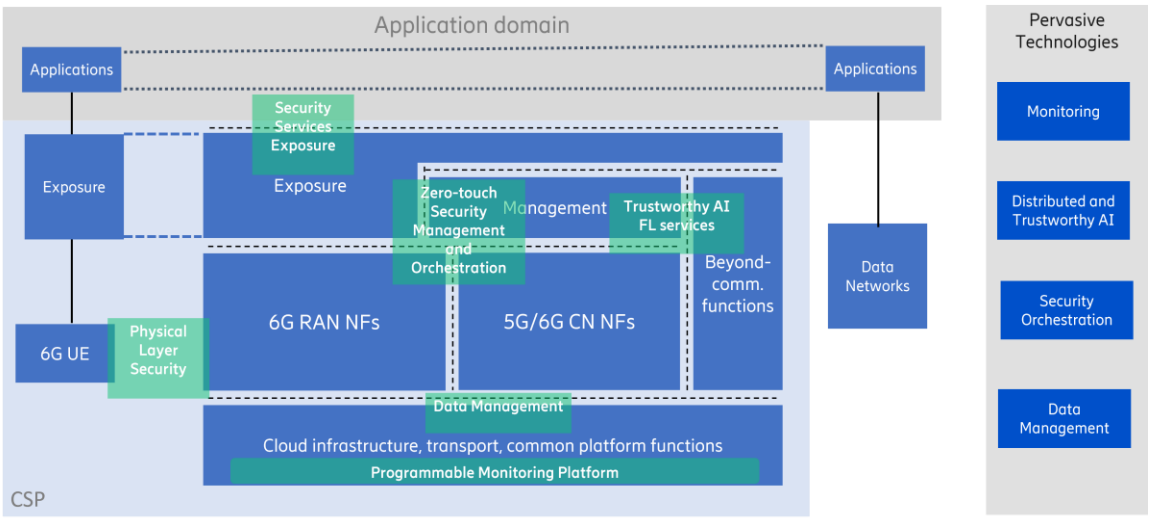

Figure 1: Deployment of ROBUST-6G macro-services enabling autonomous, explainable and secure orchestration.

Trust decisions within ROBUST-6G are enriched and validated using Explainable AI (XAI). By incorporating interpretable ML models or applying explanation frameworks to black-box classifiers, the system ensures that each security decision—whether access control, anomaly response, or reconfiguration—can be justified and traced. This explainability is crucial; if an autonomous decision were to prevent legitimate access in a high-risk scenario, the generated explanation provides the necessary justification and audit trail for operators to swiftly understand the root cause, validate the risk, or enact a controlled override. This foundation for regulatory accountability, governance

transparency, and adherence to ethical requirements like fairness is critical in federated and sensitive environments.

The architecture supports a Federated Learning (FL) framework in which distributed agents train and update security models collaboratively without sharing raw data. This preserves privacy while enabling global threat awareness. By combining FL with XAI, ROBUST-6G delivers a system that learns from diverse contexts and still provides explainable trust evaluations at the local level. Trust levels, computed based on behavior and signal integrity, inform orchestration decisions, such as function placement, access rights, and cooperation agreements.

A distinctive element of ROBUST-6G's trust management strategy is the inclusion of physical-layer security signals—such as device fingerprints, RF features, and environmental cues—into the trust computation. These signals help detect impersonation or anomalous device behavior, adding a tangible, hardware-grounded dimension to the overall security picture, making the trust assessment mechanism highly robust. This is critical for combating advanced threats where attackers may compromise software layers but cannot replicate the unique physical characteristics of a legitimate device.

To coordinate security policies and actions across domains, ROBUST-6G implements programmable interfaces and policy negotiation mechanisms. These tools allow autonomous agents to reach consensus on shared trust models or dispute resolution strategies without central control. For example, in a multi-operator environment, if one domain's local threat detection system flags a shared service function as compromised, while another domain requires that function to maintain a critical service level, a dispute resolution protocol is triggered. This protocol uses context-awareness—including the current service priority, observed threat severity, and the physical-layer trust signals of the involved entities —to negotiate a dynamic consensus, such as immediate isolation or temporary reconfiguration, ensuring minimal operational disruption. Each decision and negotiation step is logged, creating a verifiable audit trail that supports post-incident investigation and accountability.

The **iTrust6G** project addresses one of the most pressing challenges in future networked environments: how to ensure secure and trustworthy access to services and data in a world where users, devices, and services are highly mobile, dynamic, and heterogeneous. Traditional access control models based on static roles or pre-established credentials are insufficient in the 6G era, where the trustworthiness of an entity must be continuously assessed based on its behavior, context, and interaction history. To meet these demands, iTrust6G proposes a context-aware, explainable, and adaptive access control framework built upon trust computation and AI explainability.

At the core of iTrust6G's innovation is a Trust Level Computation Engine that analyses behavioral attributes—such as location patterns, usage anomalies, or deviations in communication profiles—to assign dynamic trust levels to devices and network services. These trust levels are then consumed by an AI-powered access control system that supports context-aware and explainable decision-making on trust evaluation,

access authorization and trust-aware security mitigation actions. This mechanism proactively denies access when an entity exhibits suspicious behavioral deviations (e.g., accessing sensitive data from an unusual location or time), even if static credentials are valid, addressing limitations of traditional, role-based security.

To mitigate the risk of Denial of Service (DoS) resulting from malicious profile poisoning, the Trust Level Calculation Engine employs a multi-layered data fusion approach (e.g., behavior, context, signal integrity) to prevent an anomaly in a single vector from compromising the overall score. Furthermore, every denial decision is instantly logged into the TER along with its XAI explanation, facilitating a proactive audit and dispute resolution protocol. This mechanism enables the automated or supervised reversal of erroneous or malicious lockouts, ensuring the decision process is auditable and resilient, rather than a single point of failure. Additionally, it should be noted that the TER ingests data from iTrust6G's architecture components and not directly network services and devices – providing an additional layer of protection between malicious actors and the TER's records.

As illustrated in Figure 2, the iTrust6G architecture combines behavioral monitoring with dynamic trust evaluation to enable context-aware access control decisions. Each request is analyzed in real time by AI models that are central to the Trust Assessment, detecting deviations and assigning trust levels. These trust levels then inform the final access decision. Explainable AI components ensure transparency and accountability, while a Transparent Evidence Repository logs every access decision for auditing and regulatory purposes. The TER is designed as an immutable log, providing verifiable evidentiary value for post-incident investigation and ensuring accountability. This architecture supports dynamic policy enforcement and enhances trust in highly mobile and heterogeneous environments.

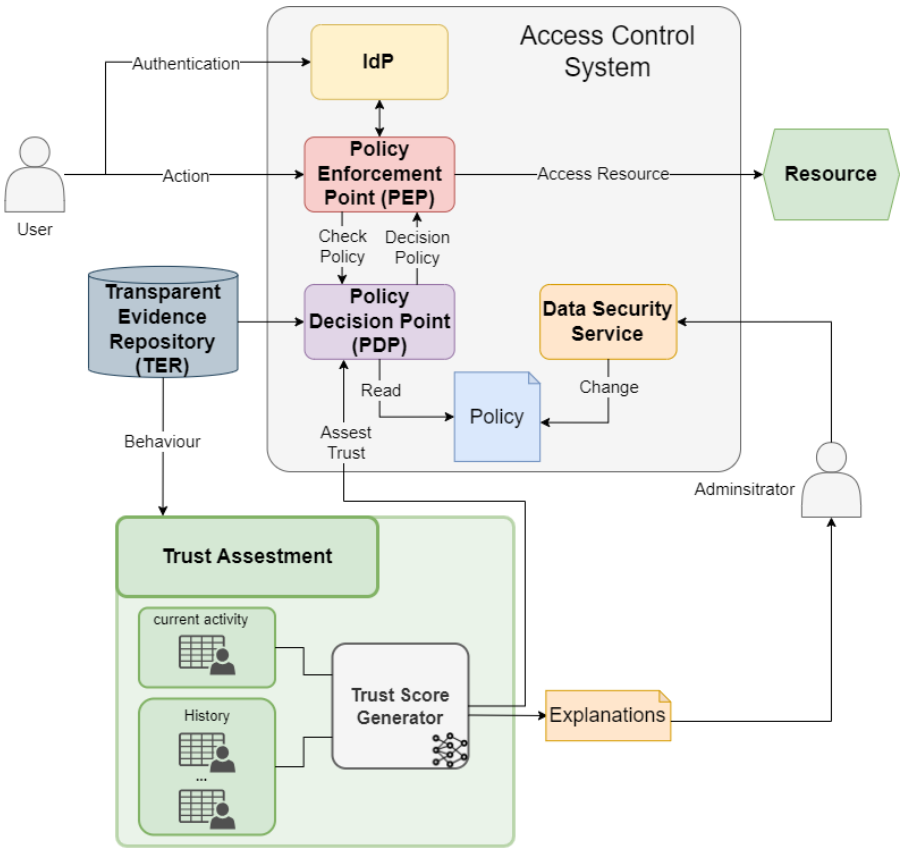

Figure 2: iTrust6G architecture integrating behavioral analysis, dynamic trust levels, and explainable access control decisions.

Crucially, the access decisions are not made in a black box. iTrust6G embeds Explainable AI (XAI) into the decision-making process to generate human-understandable justifications for each decision. This transparency is key not only for building user trust and enabling oversight but also for detecting and mitigating potential algorithmic bias, a necessary step toward ensuring fair access and operation. Whether using inherently interpretable models like decision trees or post hoc explanation tools like SHAP or LIME, the framework ensures that the logic behind granting or denying access can be understood, validated, and challenged if needed.

To ensure accountability, every access control decision and its associated reasoning are logged in a Transparent Evidence Repository (TER). This immutable log can be consulted during audits, security investigations, or compliance checks, supporting legal and operational governance. Immutability is achieved through a Security-by-Design model based on immutable ledger that includes end-to-end encryption of evidence records, pseudonymization during data ingestion, and policy-based access controls.

The confidentiality and protection of data within the Transparent Evidence Repository (TER) are cornerstones of regulatory compliance. To circumvent the risk of massive data

breaches, the TER employs a Security-by-Design model. This involves: (1) End-to-End Encryption of evidence records both at rest and in transit; (2) The use of Pseudonymization during behavioral data ingestion, ensuring that event records are not directly linkable to a user's identity without a separate, highly secured mapping key; and (3) Implementation of Policy-Based Access Controls (PBAC) that restrict access to records only to designated auditing entities for specified purposes. Responsibility for the management of this encrypted and pseudonymized data rests with the service's Security Administrator, whose activity is also auditable.

Moreover, the system supports Policy Decision Points (PDPs) that adapt access control rules in real time, based on trust scores, environmental changes, or updates in organizational policy. These PDPs can be deployed hierarchically or in a federated manner, enabling scalable policy enforcement across administrative domains.

The enabling technologies that support iTrust6G's architecture include context-collection agents, behavioral monitoring modules, and trust calibration mechanisms that adjust the sensitivity of trust levels to various inputs. These tools allow the system to operate flexibly across sectors, from industrial control systems to edge-enabled consumer applications, adapting to different privacy constraints and security requirements.

Ultimately, iTrust6G delivers a forward-looking vision of access control for 6G—one that is adaptive, explainable, context-rich, and trust-driven. It not only protects systems from unauthorized access but also establishes a culture of transparency and user sovereignty. This empowerment is achieved by granting users the right to algorithmic explanation (XAI) for every access decision and the explicit capability to audit and potentially challenge the decision logic using the immutable TER log, ensuring that 6G security mechanisms are intelligent, accountable, and human-centric.

NATWORK reconceptualizes trust and security management in 6G as a dynamic, decentralized, and adaptive process, drawing inspiration from biological systems that maintain resilience through diversity, redundancy, and continual evolution. Recognizing that 6G networks will span multiple administrative domains, support autonomous orchestration, and operate under constant change, the project proposes a security-by-design orchestration framework that emphasizes distributed trust evaluation, security compliance and intelligent defense mechanisms, and context-aware decision-making.

NATWORK secure-by-design federated orchestration will deliver capabilities in two key areas: first it will integrate security requirements of 6G slices in the deployment design and scheduling decision, thereby enabling security compliance; second, it will enable multiple edge-cloud autonomous systems (AS) to cooperate in a federated fashion on deploying a distributed 6G slice. While assuring end-to-end Quality of Service (QoS) and low latency across AS boundaries presents a key challenge, NATWORK addresses this by focusing on concatenating slice segments deployed across clusters. This ensures the end-to-end slice is provided while maintaining QoS validation at the interconnection points. Namely, to enable deployment of slice segments in local clusters with limited

information and concatenate them across clusters to provide the end-to-end slice. Cooperation is realized with minimum data sharing across ASes, to preserve AS-level confidentialities while fostering cross-AS collaboration. Here, an AS (Autonomous System) refers to a distinct administrative and operational domain, typically owned by a single service provider or enterprise, such as a Virtual Network Operator (VNO) or an Infrastructure Communication Provider (ICP). Regarding compliance, NATWORK's secure orchestration is designed to integrate deployment-time validation against a variety of frameworks, including but not limited to GDPR (for data locality and sovereignty), ISO/IEC 27000 family (for operational security requirements), and specific 3GPP/ETSI security controls relevant to Network Function Virtualization (NFV) and Mobile Edge Computing (MEC). The system allows dynamic enforcement of these diverse regulatory and organizational constraints during slice deployment and operation.

At the core of NATWORK's approach is an AI-powered Moving Target Defense (MTD) strategy. This mechanism enables the orchestrator to continuously and unpredictably reconfigure network elements —such as the placement of virtualized network functions (VNFs/CNFs) or the routing of service chains — to increase system entropy and reduce the effectiveness of targeted attacks. This continuous reconfiguration helps break the cyber kill chain, primarily by nullifying the attacker's reconnaissance or "identify" phase. By dynamically reshaping the attack surface, NATWORK promotes a defensive posture that actively anticipates threats rather than reacting to them after compromise, thereby enhancing the system's operational robustness. This makes network reconnaissance and long-term persistence by attackers significantly more costly and difficult.

A critical consideration for MTD is the integrity of service availability. NATWORK addresses this by integrating the MTD strategy with the NFV Management and Orchestration (MANO) plane, ensuring that reconfiguration actions are orchestrated, scheduled, and validated against performance and latency SLAs prior to execution. The system uses the MTD optimizer module to minimize action cost overhead and worst-case disruption time for service changes. Furthermore, recognizing that the AI decision-making component—the MTD optimizer—represents a centralized risk, its security is paramount. The optimizer itself is protected by confidential computing principles, operating within secure enclaves and utilizing runtime integrity verification for its JIT-compiled instructions. This protection makes the MTD control plane resilient against tampering, preventing the defender from becoming the liability.

As shown in Figure 3, the NATWORK MTD architecture integrates the AI-based MTD components, including the MTD optimizer and controller—with the NFV Management and orchestration plane. The architecture includes an inference engine for real-time vulnerability and trust assessment, which informs dynamic migration of services across domains. This approach supports proactive reconfiguration and adaptive orchestration policies, balancing protection, performance, and resource usage in line with the operational constraints of 6G.

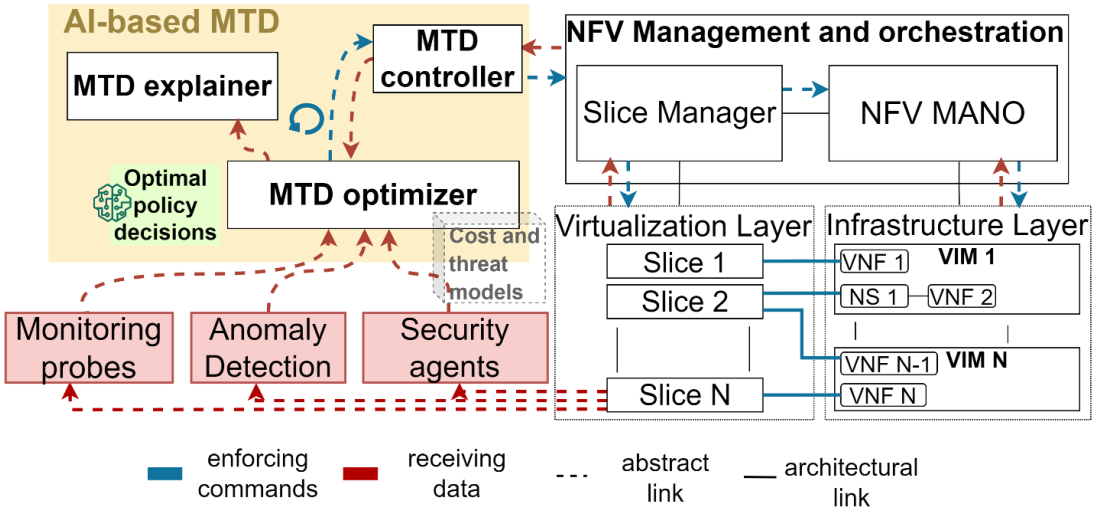

Figure 3: NATWORK MTD architecture combining service migration, inference engines, and orchestration control.

The MTD optimizer module in the NATWORK MTD architecture also introduces a new FL-based system designed to collaboratively enhance the MTD strategies employed by different VNOs (Virtual Network Operators). As illustrated in Figure 4, each VNO operates its own MTD framework instance, including a local MTD controller responsible for applying MTD actions to its own services. As a result, each VNO independently determines its defensive actions using a local multi-objective deep-Reinforcement Learning (RL) (MORL) model. One of the VNOs also acts as the Infrastructure Communication Provider (ICP), owning and managing the physical infrastructure that hosts the virtualized network slices. This VNO assumes the additional role of a central aggregator in the FL process. Since all VNOs rely on the ICP to host their slices, a baseline level of trust in the ICP is assumed to be reasonable.

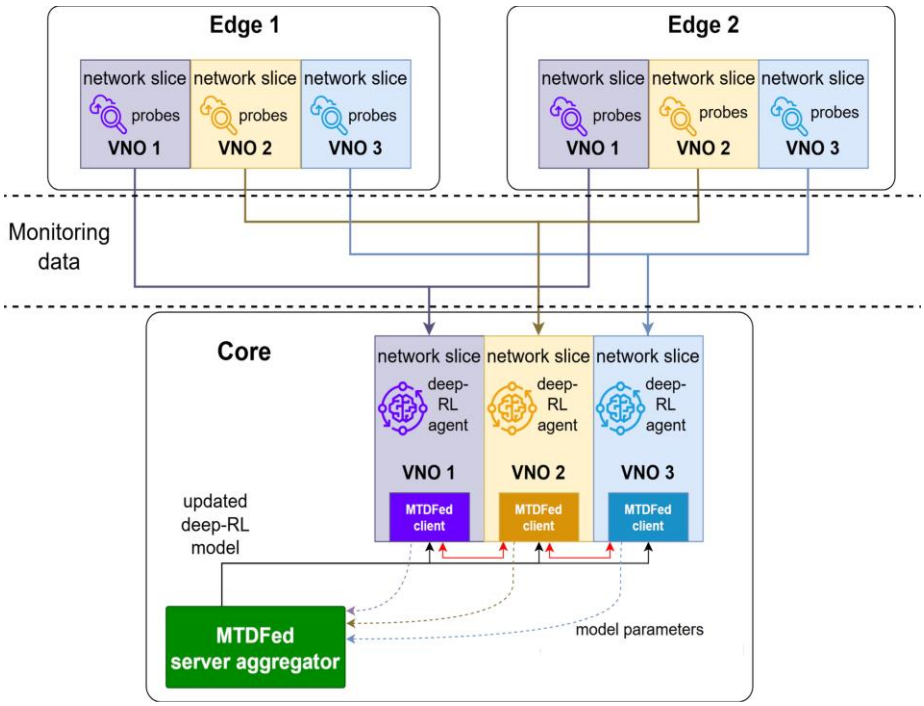

Figure 4: Architecture of the FL-based deep-RL training in NATWORK architecture.

Another distinguishing aspect of NATWORK's architecture is its support for cross-domain trust coordination, achieved through P2P CTI exchange, federated AI and policy negotiation protocols. Each domain maintains its own local models and assessments while participating in a collaborative framework that aggregates global threat intelligence and aligns security postures. Importantly, this is accomplished without requiring data centralization, thereby preserving privacy and autonomy. Trust decisions are made locally but in coordination with global indicators, enabling consistent and scalable trust management across large-scale networks.

To ensure interoperability in heterogeneous multi-operator environments, NATWORK focuses on abstracting its MTD and trust mechanisms at the orchestration layer. It does not require every peer to embed the MTD technology. Instead, NATWORK uses policy negotiation protocols and open, programmable interfaces to exchange high-level security requirements (e.g., required trust score, acceptable risk level) and Confidential Threat Intelligence (CTI) with external domains. If a partner domain does not use NATWORK's methodology, the system defaults to ensure its own service segment (slice) can meet its security SLA regardless of the partner's internal defense and uses the negotiated policies to influence external service chains at the point of interconnection. This approach allows for global trust alignment without mandating technology lock-in.

The trust orchestration loop in NATWORK is also cost-aware, meaning it does not pursue maximum security at any cost, but rather optimizes trust enforcement based on a trade-off between protection, performance, and resource usage. This multi-objective optimization ensures that security mechanisms remain efficient and do not compromise the operational integrity of the network. In scenarios where SLA compliance conflicts with resource cost optimization, NATWORK utilizes the Multi-Objective Deep Reinforcement Learning (MORL) model within the MTD optimizer. This model is explicitly trained to navigate such trade-offs by assigning weights based on real-time service priority (e.g., critical infrastructure vs. best-effort service) and the severity of the detected threat. The outcome is not simply to minimize cost, but to achieve a Pareto optimal solution that prioritizes maintaining the service level required by the SLA, even if it incurs a temporary cost overhead, ensuring business continuity remains the primary driver over immediate bottom-line reduction. The orchestrator makes informed decisions that align with both security objectives and service-level agreements.

Additionally, NATWORK enables adaptive orchestration policies, which evolve over time based on observed conditions, threat trends, and trust evolution. These policies can be re-calibrated dynamically, either automatically by the system or through operator input, ensuring that the trust model reflects the current operational environment and risk posture. This flexibility is essential for 6G, where fixed trust boundaries and static role-based controls are no longer sufficient.

6GCLOUD addresses a critical gap in the design of secure and intelligent 6G architectures by introducing the concept of the Network Digital Twin (NDT) as a foundational enabler for trust, observability, and predictive security. In contrast to traditional simulation tools or passive monitoring platforms, the NDT framework proposed by 6GCLOUD is designed to be bi-directionally integrated with live networks, capable of continuously collecting, analyzing, and acting upon real-time and historical data to enhance security and orchestration outcomes. This marks a shift toward model-driven, closed-loop security management, where network decisions are validated and optimized before deployment.

The architecture of 6GCLOUD's NDT is composed of several interconnected modules that replicate and interact with the physical network. The NDT Data Repository manages ingestion, preprocessing, and secure storage of operational data, while the NDT Engine builds simulation models to mirror the behavior, configuration, and performance of the real network. The fidelity of the simulation is crucial, and the NDT Engine is designed to model several layers of detail. This includes: (1) Network Function/Service Behavior: replicating the internal logic, resource consumption, and latency profile of individual VNFs/CNFs; (2) Traffic and Load Distribution: simulating high-volume, real-time traffic generation from thousands of UEs using probabilistic and statistical models derived from anonymized live data, including mobility patterns and diverse service requests; and (3) Infrastructure Configuration: accurately mirroring the hardware topology, resource constraints, and software configuration of the physical network segments. Critically, to address the challenge of synchronization, the NDT Engine utilizes a bi-directional telemetry stream to ensure the Digital Twin is

continuously updated with the production infrastructure's current state. This continuous synchronization is necessary for the Twin to accurately reflect real-world constraints and maintain its predictive value.

This simulation layer enables the orchestration system to test various configurations, threat responses, and policy changes in a risk-free virtual environment. Based on the simulation outcomes, the NDT provides guidance to the orchestration system on optimal deployment decisions, potential vulnerabilities, and mitigation strategies.

A central innovation in 6GCLOUD lies in its Security Management and Governance layer, which ensures that all interactions with the NDT infrastructure are controlled, auditable, and trustworthy. This includes enforcing authentication, authorization, integrity checks, and encryption for all data flows. The Lifecycle Management module ensures that digital twin instances are created, maintained, and terminated according to policy and need, avoiding unnecessary computational overhead or stale replicas. Additionally, a Performance Evaluation Module continuously assesses the reliability and confidence levels of the simulation outputs, ensuring that decisions are backed by validated predictions.

The effectiveness of the Security Management and Governance layer is directly dependent on the high level of fidelity of the NDT. Because the twin accurately replicates the network's configuration and dynamic load, the Governance layer can model and enforce compliance against frameworks (like GDPR, ISO standards, or regulatory mandates) with confidence. Specifically, it uses the simulation layer to predict the compliance impact of orchestration decisions and the propagation of threats across the network layers, providing a verifiable assurance of trustworthiness and security posture before any change is applied to the live system.

From a security perspective, the NDT acts as a trusted advisor to the orchestrator, enabling proactive threat analysis and mitigation. Before deploying any change to the live environment, the orchestrator can run multiple "what-if" scenarios in the NDT, evaluate their impact on security and performance, and deploy only those configurations that meet predefined trust and compliance thresholds. This leads to more reliable orchestration, faster reaction to evolving threats, and better resilience under dynamic conditions.

6GCLOUD's design also supports multi-layered simulations, where different NDT instances can represent various segments of the network (e.g., core, edge, RAN), and where simulations can be run in parallel to model cross-domain effects or potential conflicts. This layered view is essential for detecting cascading vulnerabilities, understanding interdependencies, and enabling security-aware orchestration in federated 6G deployments.

The work conducted by **iTrust6G** in the design context-aware access control framework for 6G network complements other lines of work in 6G trust & security management. For instance, policy-based access control supports the adaptation of network service deployment and can permit the continuous enforcement of the same regulatory constraints. iTrust6G's access control framework also serves post-incident

accountability in trust decision by collecting evidence and storing them into an immutable log. Finally, AI models involved in trust evaluations are also made explainable (XAI) to facilitate the audit ability of trust governance and promoting its transparency.

3. Privacy Awareness and Selective Confidentiality in 6G

As digital trust evolves, selective confidentiality, the dynamic and contextual protection of data, system policies, and network topology—must be a core pillar of 6G security. 6G systems must embed privacy-enhancing technologies (PETs) and architectures that guarantee control, compliance, and transparency for all sensitive assets. Two projects provide complementary approaches to privacy protection:

RIGOUROUS brings a comprehensive and forward-thinking approach to privacy awareness in 6G by embedding privacy considerations directly into the service lifecycle—from onboarding to runtime adaptation and user control. Recognizing that future networks will orchestrate services dynamically across multiple domains and administrative boundaries, the project addresses a fundamental challenge: ensuring that privacy is not an afterthought but a measurable, enforceable, and user-configurable attribute of 6G systems.

One of the core contributions of RIGOUROUS is the design and implementation of a privacy onboarding architecture that ensures privacy requirements are assessed and enforced before a service becomes operational. This is achieved through the use of privacy manifests, machine-readable descriptions that define the data protection practices, retention policies, access controls, and compliance obligations associated with a service. These manifests are analyzed during onboarding to verify whether the service aligns with organizational, legal, or user-specific privacy expectations.

Responsibility for the content of the privacy manifests is multi-tiered: the service provider is responsible for the organizational and legal sections, while the individual user interacts with a simplified, human-centric interface (PUI/AOUI) to define their personal data appetite. RIGOUROUS anticipates that adoption will be driven by a combination of: (1) Regulatory Mandate, ensuring a minimum privacy baseline; and (2) User-Centric Incentivization, where users who opt for more granular data-sharing controls receive personalized service configurations or enhanced quality of experience (QoE) based on verifiable transparency. The aim is to shift from mandated blanket policies to negotiated privacy agreements that reward user engagement and control.

The orchestration process in RIGOUROUS is enhanced with Security Orchestration, Automation and Response (SOAR) loops, which monitor service behavior during execution and compare it to declared privacy intentions. If discrepancies or violations are detected—such as unauthorized data sharing or excessive data collection—the SOAR loop triggers automatic responses, such as reconfiguring service chains, limiting access, or halting deployments. This dynamic adaptation ensures that privacy is continuously enforced, even as the service environment changes.

While actions like halting deployments may seem commercially undesirable for general use cases, the RIGOUROUS framework is primarily targeted at mission-critical systems and vertical industries with stringent security requirements—such as industrial control systems (ICS), healthcare, and sensitive government infrastructure.

For these high-assurance environments, compliance with data sovereignty, regulatory adherence, and integrity preservation outweighs commercial availability concerns. For less critical services, the SOAR response can be configured to less disruptive actions, such as isolating the service segment or triggering warnings, maintaining flexibility across different trust models.

As depicted in Figure 5**,** the RIGOUROUS framework integrates privacy enforcement throughout the orchestration lifecycle. Privacy manifests are evaluated during onboarding to verify compliance with policy and user expectations. SOAR loops continuously monitor runtime behavior and respond to violations automatically. Users can control and understand privacy implications via a dedicated interface that communicates risk levels and suggests compliant alternatives. This architecture ensures measurable, adaptive, and user-centric privacy in complex multi-domain 6G environments. This interaction is handled by the Human Controllable Privacy UI (PUI) and the Application Onboarding UI (AOUI), which are the user-facing gateways in Figure 5. The PUI abstracts complex policy data into a quantifiable privacy risk score (e.g., a color-coded or numerical index) and suggests compliant alternatives (e.g., "Use this alternative network slice which has a lower privacy impact score"). This design emphasizes clear, actionable communication of risk rather than relying on complex API data exchange, making the system accessible to non-expert users and facilitating genuine privacy awareness.

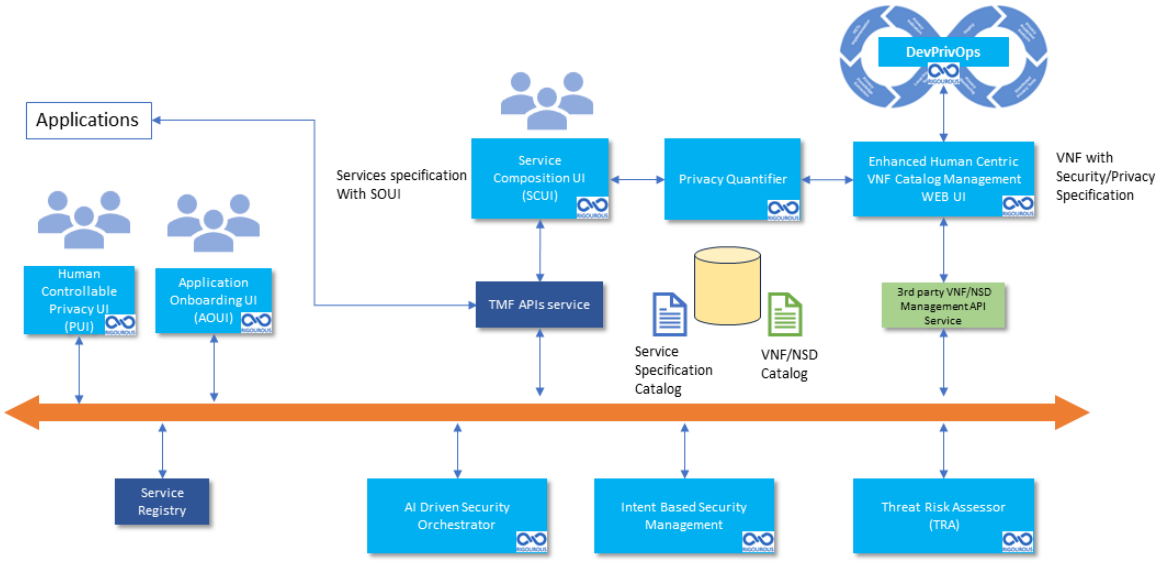

Figure 5: Privacy-aware orchestration framework in RIGOUROUS with onboarding, runtime monitoring, and user interfaces.

The project also introduces formal privacy metrics and quantification models, which allow the orchestration engine to assign a privacy score to each service or configuration. These scores can be used to compare candidate services and guide orchestration decisions, ensuring that the most privacy-preserving option is selected whenever possible. Importantly, these metrics are not based solely on static policy

analysis but also on empirical data derived from runtime monitoring and historical performance.

Another key innovation is the integration of Security Testing Digital Twins (STDTs) into the privacy assurance workflow. These sandboxed environments replicate real network and service behaviors, allowing orchestration systems to simulate data handling practices, evaluate potential privacy breaches, and test the impact of policy enforcement mechanisms before applying changes to the live environment. To ensure the STDT itself maintains the privacy-protecting goal, all simulation activities use synthetic or anonymized data to prevent any compromise of sensitive information. This predictive capability adds an additional layer of assurance and helps minimize the risk of privacy violations in production networks.

NATWORK addresses privacy in 6G through a deeply integrated architectural and operational perspective, recognizing that privacy cannot rely solely on static policies or perimeter-based defenses. Instead, the project proposes a model where privacy becomes an active component of orchestration logic and network decision-making, leveraging distributed, collaborative, and adaptive techniques. In this vision, privacy is treated as a dynamic, contextual property to be optimized alongside operational factors like latency and energy efficiency.

It is critical to clarify that NATWORK's orchestration-centric approach primarily enhances infrastructure privacy and collaborative intelligence privacy. This includes protecting service topologies, operator policies, and the confidentiality of data assets used for AI model training (e.g., Federated Learning inputs). It does not, by design, address threats specific to the User Equipment (UE) or Application Layer, such as Pegasus-type spyware or Graphite attacks where content confidentiality is breached after decryption on the device. For these threats, end-to-end encryption and device security measures remain essential, while NATWORK focuses on preventing the leakage of sensitive network and operational data at scale.

A foundational element of NATWORK's approach is its intensive use of Federated Learning (FL) to train orchestration and security models without sharing raw or sensitive data between domains. Since this protection is inherent to the FL training paradigm, it provides a crucial layer of data confidentiality management and, where personal data are involved, enables privacy preservation with minimal architectural overhead. Through FL, multiple entities can jointly train AI models while keeping their data local, thus supporting regulatory compliance. This approach is essential in 6G, where network functions and user data will be highly distributed across diverse operators, jurisdictions, and trust boundaries.

To further reinforce data protection, NATWORK integrates advanced Privacy Enhancing Technologies (PETs) such as Differential Privacy (DP) and Secure Multiparty Computation (SMC). These mechanisms ensure that the confidentiality of the training process is preserved: DP mitigates data leakage risks during model aggregation, while SMC protects the sensitive intermediate outputs (e.g., model gradients) shared among

domains. This rigorous approach enables confidential collaborative intelligence across potentially untrusted domains without compromising data sovereignty.

Operationally, NATWORK embeds privacy-aware logic into orchestration decisions. For instance, when selecting a deployment location for a network function or service, the orchestrator evaluates not only technical parameters like latency or resource availability but also privacy implications related to jurisdiction, historical trust performance of the node, and the sensitivity of the data involved. This is achieved through the integration of distributed trust scores and contextual privacy profiles that enable real-time orchestration decisions aligned with privacy goals.

A cornerstone of NATWORK's confidential computing strategy is the use of Federated Learning (FL), which enables the training of AI models across multiple domains without the need to centralize data. This approach operates on the hypothesis that each domain trusts its local execution infrastructure (where data processing occurs) but must protect its internal data from the collaborating peer domains, whose behavior is deemed untrustworthy. By keeping data local and sharing only model updates, NATWORK ensures that sensitive information never leaves the trusted perimeter of each domain. However, recognizing that even metadata can leak information, the project strengthens FL with Secure Multiparty Computation (SMC) and Differential Privacy techniques, making collaborative learning secure by default. This architecture allows NATWORK to establish confidential collaborative intelligence, where security-critical orchestration and trust decisions are made based on shared insights without exposing internal data or policies.

NATWORK further supports privacy compliance when sharing CTI data across domains, in which the amount of exposed information is adapted to the confidentiality needs of the publishing domain and the potential risk of the CTI.

The architecture also supports cross-domain privacy policy negotiation mechanisms, allowing domains to exchange privacy requirements and capabilities before establishing collaboration. This removes assumptions about implicit trust and replaces them with explicit, verifiable, and observable decision processes, increasing transparency and accountability.

The challenge of communicating the "privacy-protected" status to a roaming UE is acknowledged. While NATWORK's core focus is on inter-domain privacy guarantees, the system enables the host network to expose the privacy baseline offered (e.g., data locality, data retention policy, and compliance level) through standardized 6G-API interfaces. This allows the UE or the user's home network to dynamically assess the risk of the host domain based on its verified privacy posture. Although full disclosure of the orchestration logic is impractical, the negotiation mechanisms ensure that the local domain's orchestration decisions satisfy the privacy constraints demanded by the home network, establishing a verifiable, minimum guarantee for the user even when roaming.

4. Confidential computing in 6G

"As computation moves to the edge and services are deployed across increasingly heterogeneous and potentially untrusted platforms, Confidential Computing becomes essential in ensuring that sensitive data and workloads remain protected—even during processing. *This involves distinguishing Confidential Computing (protection of data in process from the host) from generalized Confidentiality (protection of data in transit/at rest).* This capability is critical not only for general data confidentiality but also as a core mechanism supporting user privacy, particularly when processing highly sensitive personal data within untrusted domains. Projects across the SNS portfolio are addressing this with architectural innovations."

ELASTIC contributes to the vision of confidential computing in 6G by developing an architecture that enables secure, privacy-preserving, and verifiable execution of services across highly distributed and heterogeneous environments. As 6G networks push computing further toward the edge and embrace a serverless, composable paradigm, the challenge of maintaining data confidentiality and trust in untrusted or semi-trusted infrastructure becomes central. This necessitates an "ultra-paranoid" security posture that addresses supply chain integrity, trusted execution, and fine-grained network security, moving beyond traditional perimeter defenses to guarantee data protection throughout its entire processing lifecycle across heterogeneous domains. ELASTIC addresses this challenge by combining lightweight virtualization technologies, attestation mechanisms, and decentralized service orchestration to create a trustworthy foundation for next-generation network services. To be precise, the confidential computing capability is realized through the integration of these techniques with Trusted Execution Environments (TEEs) and cryptographic CPU hardware extensions, which are essential for data protection during processing.

At the heart of ELASTIC's architecture is the use of WebAssembly (Wasm). Wasm is a portable binary instruction format that provides a sandboxed, low-footprint runtime, enabling services (compiled from over 40 programming languages) to be deployed flexibly across the continuum. Wasm modules are deployed within secure execution environments (SEEs) enabled by TEEs, thus enabling verifiable, confidential processing even in untrusted locations. The framework also introduces attestation and integrity verification mechanisms to validate the trustworthiness of execution environments before and during service execution.

Figure 6 illustrates the ELASTIC modular architecture, which is constructed from four interconnected technology blocks (Wasm, FaaS, Confidential Computing, and eBPF/XDP) to support confidential and privacy-preserving orchestration across heterogeneous and distributed 6G environments. The architecture leverages Wasm — a portable binary format requiring a Wasm runtime execution environment— as the foundational component, offering a lightweight and portable alternative to traditional containers. It is complemented by TEE-based secure enclaves to ensure trustworthy execution, and eBPF hooks for fine-grained observability and dynamic policy

enforcement. The inclusion of a Federated Learning Toolbox enables cross-domain model training without data centralization, thus enhancing privacy. Additionally, serverless agents and attestation components enable dynamic code trust validation before and during execution, supporting zero-touch, secure orchestration across the cloud-edge-device continuum.

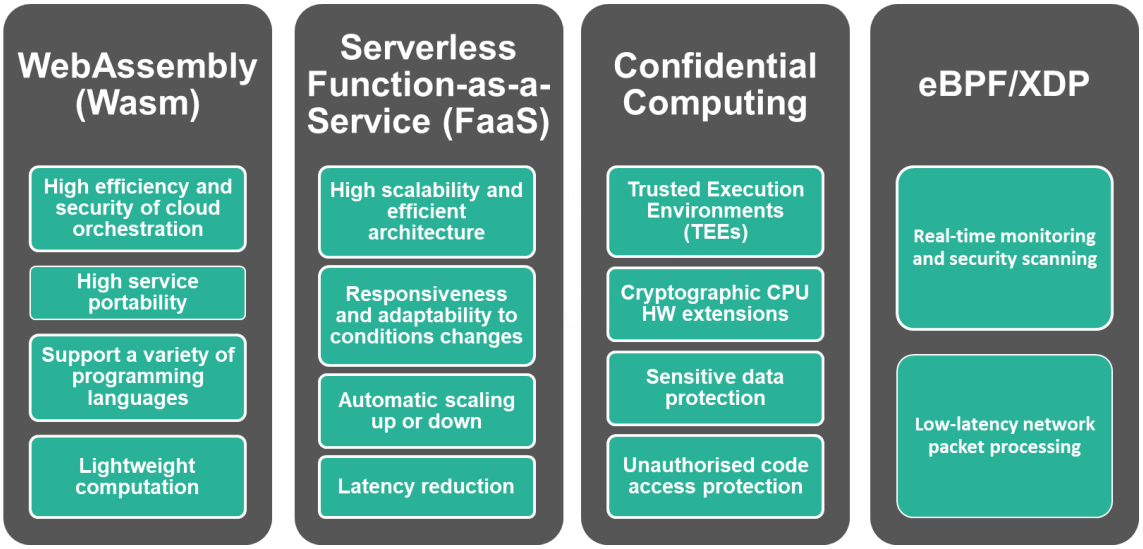

Figure 6: ELASTIC's modular architecture for confidential service orchestration using Wasm, attestation, and eBPF.

To further enhance trust, ELASTIC incorporates remote attestation mechanisms that verify the integrity and provenance of Wasm modules before execution. These attestation processes confirm that the code running on the device has not been tampered with and originates from a trusted source, enabling orchestrators and service providers to deploy critical logic with verifiable execution integrity even on infrastructure that may not be fully trusted. This forms the basis for building confidential execution environments across a distributed, federated network.

In addition to secure execution, ELASTIC introduces mechanisms for confidential orchestration that tightly protect service deployment and management metadata. This involves securing the orchestration control plane and encrypting communication to prevent sensitive deployment strategies or service topologies from leaking. This level of confidentiality is essential in 6G scenarios involving vertical industries, critical infrastructure, or cross-operator collaboration, where both data and orchestration logic must be protected.

The architecture also supports mobility-aware and context-sensitive service migration, allowing Wasm-based agents to move between nodes in response to mobility events or load balancing needs—without exposing sensitive state or compromising security. During migration, attestation is re-evaluated, and secure context transfer mechanisms ensure that sensitive state remains protected throughout the lifecycle of the service.

ELASTIC enforces a data-centric confidentiality model by integrating fine-grained controls into the orchestration layer. This model dictates what data can be accessed, how it is processed, and where the computation is allowed to occur. A key enabler for this is the ability to automatically generate confidentiality- and privacy-enhancing policies based on service descriptors, deployment context, and user preferences, thus aligning service behavior with protection constraints throughout the service lifecycle.

NATWORK contributes to the field of confidential computing in 6G by embedding security guarantees directly into the orchestration and service deployment process, especially across federated and multi-stakeholder environments. A distinctive technical contribution of NATWORK lies in its identification and development of WebAssembly as a foundational technology for secure 6G service deployment. While not providing hardware isolation itself, NATWORK leverages Wasm's core security features—specifically its built-in sandbox and isolation-by-design—to enforce software-level confidentiality and address specific security challenges where TEEs (Trusted Execution Environments) are unavailable or impractical, particularly on diverse, resource-constrained edge devices. The project couples Wasm's isolation capabilities with software-based attestation mechanisms to create a verifiably secure execution environment that aligns with the broader goals of confidential computing.

As 6G networks push computation toward the edge and embrace serverless, highly mobile service paradigms, traditional containerization approaches face limitations in terms of startup latency and resource overhead. Wasm offers advantages in these areas while sharing a similar level of universal portability with modern containers. Wasm technology emerged from browser security requirements but has evolved into a compelling alternative to traditional containerization for 6G services. Wasm extends seamlessly to User Equipment (UE), enabling true continuum deployment without vendor lock-ins, featuring ultra-fast startup with zero penalty for service instantiation, built-in sandbox environment providing process isolation by design, type-controlled execution preventing buffer overflow attacks, and lower-level instruction sets that are harder to comprehend (but not reverse) than plain JavaScript.

However, NATWORK's analysis reveals critical security challenges specific to Wasm in 6G contexts, including code tampering vulnerabilities through privilege escalation or direct memory introspection, as Wasm bytecode lacks the write-execute protection of compiled binaries, and Just-In-Time (JIT) spraying attacks where malicious actors manipulate Just-In-Time compilation to generate exploitable native code snippets. To address these challenges, NATWORK develops a runtime integrity verification system that operates beyond traditional authentication mechanisms. While existing approaches focus on pre-execution verification, NATWORK introduces continuous runtime monitoring to detect tampering during execution. Recognizing the typical performance overhead associated with such continuous integrity verification, NATWORK’s system is optimized to balance the three dimensions of security, performance, and sustainability. The technical approach involves modifying the WASMTIME open-source runtime to embed a secondary verification thread that monitors the integrity of JIT-compiled machine instructions, performing continuous integrity checks against known-good baselines and triggering automated responses

when tampering is detected. Our solution also tackles with overhead caused by periodic runtime integrity verification on the Wasm modules by restricting the measuring thread allocated resources using Linux's cgroups.

As illustrated in Figure 7, the NATWORK WebAssembly module integrity workflow will be integrated with D-MUTRA, derived from the DESIRE-6G project. D-MUTRA is a blockchain-based mutual remote attestation framework that provides a platform-agnostic, zero-touch solution for remote attestation. The blockchain orchestrates the various remote attestations, distributing one-time verification tasks across different nodes and logging the attestation results.

In this implementation, verifications are performed by modified WASMTIME runtimes, which are accessed through the blockchain. This framework not only enables remote attestation of Wasm payloads but also continuously checks their integrity during execution. This architecture addresses the need for lightweight, cross-platform confidential computing in dynamic, multi-domain 6G infrastructures.

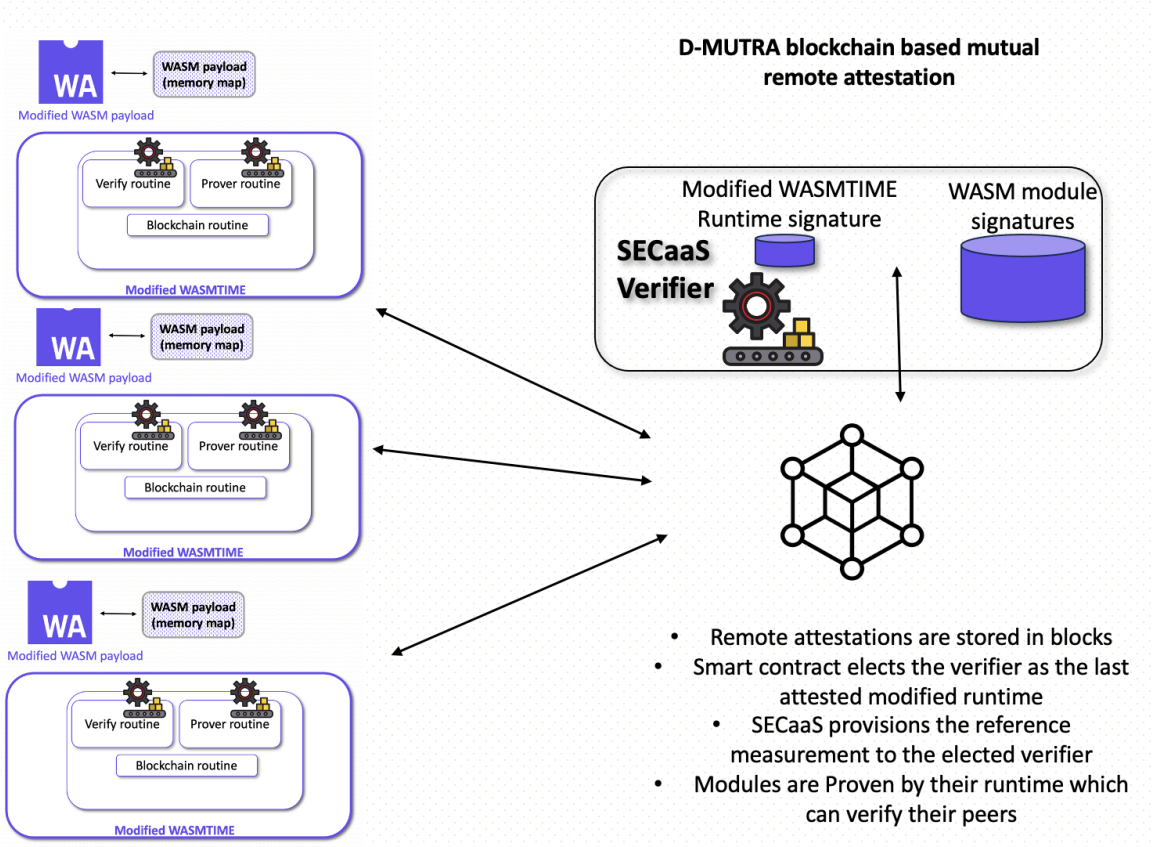

Figure 7: NATWORK Wasm module blockchain based remote attestation using D-MUTRA.

In addition to Wasm module integrity, NATWORK will also address confidentiality and availability by further modifying the WASMTIME runtime. Beyond Wasm payloads, NATWORK explores how containerized workloads and their orchestration platforms (e.g., Kubernetes) can be leveraged to accelerate the adoption of D-MUTRA—most

notably by enabling a fully automated, zero-touch setup in which containers are continuously verified during execution. Since Wasm runtimes are frequently deployed within containers, our work in both areas will naturally converge.

In addition to data privacy, NATWORK incorporates mechanisms to protect orchestration logic and deployment metadata. The project introduces orchestration workflows that operate under constrained visibility, ensuring that no single domain has a full view of the system unless explicitly authorized. This limits the exposure of sensitive policies, resource usage patterns, and network topologies, all of which could be exploited if leaked to adversaries or untrusted intermediaries. NATWORK manages the trade-off inherent in constrained visibility by using policy negotiation protocols to share resource commitment summaries without exposing raw usage patterns, thereby mitigating the risk of denial of service by resource exhaustion.

Another key aspect of NATWORK's confidential computing model is its ability to dynamically evaluate and enforce context-aware confidentiality constraints. Orchestration decisions are driven by privacy and trust policies that consider data sensitivity, node capabilities, jurisdictional requirements, and threat levels. Services are deployed only in environments that can meet the required confidentiality level, and migration or reconfiguration is triggered automatically when these conditions are no longer satisfied.

Through this multi-layered and adaptive approach, NATWORK provides a collaborative, federated model of confidentiality that goes beyond the isolation of execution environments. This framework ensures that both data protection and policy confidentiality are maintained across dynamic, heterogeneous networks. This capability is essential in the context of 6G, where sensitive workloads are distributed across organizational and national boundaries, demanding confidentiality to be preserved across administrative domains.

ROBUST-6G brings a multidimensional perspective to data protection in 6G by integrating privacy-preserving mechanisms, trustworthy federated intelligence, and secure infrastructure capabilities. Each of these components is designed to cover different untrust scenarios: Secure infrastructure capabilities (TEEs) protect the workload from the underlying platform; privacy-preserving mechanisms (DP) safeguard data from providers; and federated intelligence addresses untrustworthiness among collaboration partners. ROBUST-6G addresses the critical need to protect sensitive data, inference models, and decision-making processes during execution and learning, especially where dedicated hardware enclaves are not available.

One of the central contributions of ROBUST-6G in this area is the development of a federated learning framework that not only avoids centralizing raw data but also embeds confidentiality guarantees throughout the lifecycle of training and inference. This is achieved by integrating differential privacy to mitigate data leakage risks, as well as mechanisms to obfuscate model updates when collaborating across less-trusted domains. The system ensures that the knowledge generated by each participant is protected while still enabling the aggregation of global security intelligence.

Beyond data protection, ROBUST-6G addresses the confidentiality of model execution and decision-making. Given that security and trust decisions may involve analyzing sensitive contextual signals—such as device behavior, location data, or user interaction patterns—the project deploys explainable AI (XAI) techniques that, to fulfill the definition of true Confidential Computing, operate within secure enclaves or trusted execution environments (TEEs) when available. These environments prevent external observation or tampering while allowing inference operations to remain interpretable and verifiable.

Another unique aspect of ROBUST-6G's confidential computing strategy is its emphasis on macroservice modularity. Each security macroservice can function autonomously within isolated domains, ensuring that critical policy enforcement, monitoring, or response capabilities can run locally under confidential conditions. These services exchange only abstracted and anonymized signals with other macroservices, preserving confidentiality even as they cooperate for threat detection or response orchestration.

The project also considers the challenge of orchestration metadata confidentiality, recognizing that exposure of deployment strategies, trust policies, or resource allocation decisions can itself create security risks. To address this, ROBUST-6G employs secure APIs and encrypted policy negotiation protocols that protect orchestration logic from adversarial observation or manipulation, especially in multi-operator environments. This security relies on the verifiable service identities of the communicating autonomous agents, typically established through mutual authentication mechanisms like secure digital certificates.

Finally, ROBUST-6G introduces support for context-sensitive policy enforcement, where privacy and confidentiality requirements evolve alongside changes in risk posture, network conditions, or service context. Policies may dictate, for instance, that a service is only executed on hardware supporting attestation, or that data is processed only in domains certified under specific compliance frameworks. The orchestration engine ensures that these conditions are continuously evaluated and enforced, adapting service placement or behavior as needed.

SUNSET-6G explores sustainable security by focusing on:

- Reducing the energy consumption of continuous security services, particularly in AI-based anomaly detection, orchestration, and encryption.
- Designing security architectures that balance protection with environmental efficiency, avoiding unnecessary computation and communication overhead.
- Introducing metrics that link security assurance to sustainability indicators, ensuring security solutions are not only technically sound but also ecologically responsible.

To flesh out the concept of linking security and sustainability, SUNSET-6G is developing a framework centered on two types of preliminary metrics: (1) Security Cost Overhead

(SCO), measured as the ratio of energy (Joules) consumed by security functions (e.g., continuous AI-based monitoring, encryption, MTD actions) relative to the total network power consumption. This SCO metric provides a direct measure of the resource-conscious design. (2) Assurance-per-Watt (ApW), which measures the security level achieved (Assurance Score) against the energy cost (Watt) of the defense mechanism. By minimizing the SCO and maximizing ApW, the project ensures that robust protection does not compromise environmental efficiency.

While still in early phases, SUNSET-6G contributes to the growing awareness that security and sustainability must co-evolve in future network designs, establishing a new dimension of trade-off in 6G systems.

This approach contrasts with the Level of Trustworthiness (LoTw) KVI introduced by SAFE-6G. While LoTw is a weighted sum focused on assurance dimensions (Safety, Security, Privacy, Resilience, Reliability), SUNSET-6G proposes a necessary pre-condition: that the security mechanisms supporting LoTw must first be validated for sustainable operation. The goal is to ensure that achieving high trustworthiness (high LoTw score) is not ecologically prohibitive, effectively extending the KVI framework to include a measurable environmental cost dimension.

5. KPI/KVI for Security in 6G

As 6G networks become increasingly autonomous, distributed, and dynamic, the definition and tracking of Key Performance Indicators (KPIs) and Key Value Indicators (KVIs) for security become indispensable. Unlike previous generations, 6G requires security to be:

- **Measurable** in real time,
- **Aligned with service performance objectives**, and
- **Transparent and verifiable** for both technical stakeholders and regulatory bodies.

Projects across the SNS programme have begun to propose diverse and advanced KPIs/KVIs tailored to their security innovations.

The initial set of metrics proposed by these projects primarily focuses on performance assurance (e.g., time-to-detect/mitigate, latency, overhead) for security functions. While these operational KPIs are indispensable, the sheer volume and diversity of these performance metrics highlight the pressing need for a unified framework that connects these granular measurements to a higher-level, business and societal value indicator.

Together, these metrics reveal the need for a broader security KVI framework in 6G that:

- Connects security functions with **operational KPIs** (latency, availability, cost),
- Exposes **sustainability-security trade-offs** and the perceived environmental responsibility of secure networking,
- Expresses **societal resiliency** (index / score),
- Enables **continuous, real-time trustworthiness assessment and benchmarking** in orchestration and AI, addressing the inherent difficulty in precisely measuring complex metrics such as protection gain, explainability, and orchestration success.
- Facilitates **compliance visibility** for regulators and providers, and
- Supports **real-time adaptive feedback loops** in secure network management.

SAFE-6G project introduces an innovative approach to evaluating and ensuring the alignment of 6G technologies with societal, environmental, and economic priorities through the adoption of KVIs. SAFE-6G is a key initiative addressing the need for a unified KVI framework. Its approach directly aligns with the diverse set of granular KPIs/KVIs proposed by other SNS projects—such as the time-to-mitigate (HORSE), protection gain (NATWORK), and explainability scores (iTrust6G)—by providing a

methodology to aggregate them into a single, verifiable score. This methodology ensures the translation of operational security performance into a meaningful, system-wide measure of value. A key feature of the SAFE-6G framework is the integration of Trustworthiness metric, which is used to evaluate the dependability and security alignment of the system with the needs of the user/tenant through an intelligent chatbot capable of capturing the user-intent. This approach addresses the inherent complexity of translating granular metrics (like privacy, security, resilience, reliability, safety) into an aggregated, operational, and verifiable trustworthiness score, thereby meeting the challenge of KVI measurability.

The Key Value (KV) Trust, selected as the more relevant with the SAFE-6G scope and activities focuses on identifying and quantifying how trust-related factors—safety (avoiding catastrophic failure or harm), security (protecting from intentional attacks), privacy (ensuring control over personal data), reliability (consistent and error-free operation over time), and resilience (the ability to recover from failure)—influence both system and user interactions. By analyzing the enablers and barriers that shape trust, this KV highlights how these elements enhance the overall value of trust and trustworthiness across multiple domains. The SAFE-6G consortium distinguishes between "Trust," understood as an attitude that a tenant has toward a 6G system, and "Trustworthiness," a measurable property of the system that fosters trust in its users. The more trustworthy a 6G system is, the higher the perceived trust level by the tenant will be. Building on this, the Key Value Indicator (KVI) Trustworthiness represents a paradigm shift from traditional security-focused thinking toward a broader concept of native trustworthiness encompassing five core dimensions: Safety, Security, Privacy, Resilience, and Reliability. The trustworthiness approach presented in SAFE-6G is also supported by the UNITY-6G project, which integrates AI-native mechanisms with DLT-based transparency, semantic communications, and digital twinning to enhance security, reliability, and explainability. UNITY-6G's conflict-resolution and trustworthy-AI components further reinforce SAFE-6G's principles by ensuring fair, safe, and transparent resource allocation across tenants. Each dimension is implemented through a dedicated Trust Function operating across the application plane, the 6G core, and the edge/cloud continuum. These Trust Functions continuously assess contextual information, collaborate with a Cognitive Coordinator to reason intelligently, and perform targeted actions to ensure that the system remains in a trustworthy state. Each Trust Function outputs a score, denoted as LoTFj, representing its contribution to overall trustworthiness. The Cognitive Coordinator aggregates these using a weighted formulation to compute the Level of Trustworthiness (LoTw) as follows:

$$\text{LoTw} = \sum_{j=1}^{N} W_{LoTFj} \cdot LoTF_j$$

subject to the constraint that the sum of all weights assigned to the Trust Functions equals one. Thus, trustworthiness is not only quantifiable and explainable but also actionable in real time, enabling a dynamic and human-centric approach to trust. This mechanism ensures that trustworthiness is dynamically calculated, context-aware, and operationalized, fulfilling the SAFE-6G objective of measurable and enforceable trust in 6G systems.

Future work in the Security Working Group should focus on harmonizing these KPIs/KVIs into a shared model that can be used across testbeds, trials, and vertical applications.

6. Standardization activities

Security-related innovations developed across the SNS projects have clear implications for standardization and alignment with ongoing initiatives in international bodies. These contributions span multiple domains including orchestration, AI, privacy, and trust.

HORSE aligns its policy-aware orchestration framework with ETSI ZSM (Zero-touch Service Management) and ENI (Experiential Network Intelligence), while providing feedback on Digital Twin integration into security workflows, relevant for future specifications.

RIGOUROUS contributes to ETSI ZSM, ENI, and 3GPP SA5 by proposing mechanisms for secure service onboarding, zero-trust orchestration, and continuous risk assessment using Security Testing Digital Twins. Its work also aligns with NIST guidelines for automated trust quantification and AI governance.

ELASTIC builds on WebAssembly and contributes to the W3C WebAssembly community by proposing extensions for secure deployment in edge environments. It also links ETSI MEC and NFV standardization activities through its confidential orchestration interfaces.

ROBUST-6G provides architectural input to NIST's AI Risk Management Framework and ETSI ZSM, particularly regarding explainability and physical-layer security. Its work on trustworthy and explainable federated AI aligns with future governance recommendations for autonomous systems.

iTrust6G maps directly to ETSI TS 104 224 (Trustworthy AI lifecycle) and supports the ISO/IEC 27000 family—specifically aligning with ISO/IEC 27002 controls for access and identity management—by proposing explainable access control architectures and dynamic trust metrics. It complements AI-related initiatives under IEEE and ISO/IEC JTC 1/SC 42.

NATWORK contributes to ETSI NFV and MEC, especially in dynamic service relocation and secure orchestration using MTD. Its federated AI approach also engages with ENISA and emerging policy frameworks on cross-domain security and data protection.

Coordinated input to standardization bodies will be essential to consolidate these advancements and ensure their integration into the foundational technologies of 6G.

7. Conclusions

The collective work carried out across the SNS projects provides a rich and forward-looking perspective on the security architecture of 6G, establishing a fundamental transition from static, reactive security to a proactive, intrinsic, and data-centric paradigm.

The collective efforts consolidate three essential architectural principles that underpin a trustworthy 6G ecosystem:

First, 6G Security Must Be Predictive and Modelled with Network Digital Twins (NDTs). The NDT and the Security Testing Digital Twin (STDT) are established as foundational tools for real-time risk simulation, validation, and optimization of security decisions before any live deployment. This capability ensures that countermeasures are verifiable and that security is integrated into the service lifecycle.

Second, Trust is a Dynamic, Explainable, and Behavior-Based Level. Traditional trust models are insufficient for the dynamic nature of 6G. The work drives a model where an entity's trust level is dynamically assigned based on contextual behavior analysis. This adaptive trust is coupled with Explainable AI (XAI), which ensures that autonomous access control and threat mitigation decisions are transparent, auditable, and traceable within the Transparent Evidence Repository (TER).

Third, Confidentiality Extends to the Edge and Is Achieved Through Lightweight Isolation and PETs. Protecting data and workloads while processing across the cloud-edge continuum is achieved through architectural innovation. Projects promote WebAssembly (Wasm) as a lightweight, sandboxed alternative to containers, combined with Trusted Execution Environments (TEE) and Remote Attestation for verifiable execution integrity in resource-constrained environments. Furthermore, Confidential Collaborative Intelligence is made possible through the systematic use of Privacy-Enhancing Technologies (PETs) such as Federated Learning (FL), Differential Privacy (DP), and Secure Multiparty Computation (SMC), which secure shared insights without exposing raw data.

These architectural shifts are implemented through concrete advancements across core security dimensions. This includes the implementation of Zero-touch Orchestration, the deployment of Moving Target Defense (MTD) mechanisms, and Privacy-aware Service Onboarding via Privacy Manifests.

Finally, the introduction of Key Performance Indicators (KPIs) and Key Value Indicators (KVIs) tailored to security and trustworthiness, including pioneering metrics like Assurance-per-Watt (ApW)—signals a necessary shift from static compliance models to measurable, explainable, and sustainable performance-based security. Coordinated contributions to standardization bodies like ETSI, 3GPP, ITU-T, NIST, and W3C are essential to scale these innovations and ensure their integration into the foundational technologies of 6G.

References

[1] https://horse-6g.eu/
[2] https://rigourous.eu/
[3] https://elasticproject.eu/
[4] https://robust-6g.eu/
[5] https://www.sns-itrust6g.com/
[6] https://www.6g-cloud.eu/
[7] https://natwork-project.eu/
[8] https://cris.vtt.fi/en/projects/sustainable-network-security-tech-for-6g
[9] L. Kastner, «On the Relation of Trust and Explainability: Why to Engineer for Trustworthiness,» de IEEE 29th International Requirements Engineering Conference Workshops (REW), Notre Dame, IN, USA, 2021

List of Editors

Editor Name	Institution	Country	Project
Antonio Skarmeta	University of Murcia	Spain	HE SNS RIGOROUS
Noelia Pérez Palma	University of Murcia	Spain	HE SNS RIGOROUS
Dhouha Ayed	Thales	France	HE SNS ELASTIC

List of Contributors

Contributor Name	Institution	Country	Project
Antonio Fernando Skármeta Gómez	University of Murcia	Spain	SNS RIGOROUS
Noelia Pérez Palma	University of Murcia	Spain	SNS RIGOROUS
Dhouha Ayed	Thales	France	SNS ELASTIC
Rhys Miller	Gigasys Solutions		SNS iTrust6G
Maxime Compastié	i2CAT Foundation	Spain	SNS iTrust6G
Juan Manuel Montez López	UC3M	Spain	SNS iTrust6G
Lucia Cabanillas Rodrigez	TID	Spain	SNS iTrust6G
Diego R. Lopez	TID	Spain	SNS iTrust6G
Pablo Serrano Yañez-Mingot	UC3M	Spain	SNS iTrust6G
Wissem Soussi	ZHAW	Switzerland	SNS NATWORK
Gürkan Gür	ZHAW	Switzerland	SNS NATWORK
Vincent Lefebvre	Solidshield	France	SNS NATWORK
José María Jorquera Valero	University of Murcia	Spain	SNS ROBUST-6G
Manuel Gil Pérez	University of Murcia	Spain	SNS ROBUST-6G
Gregorio Martínez Pérez	University of Murcia	Spain	SNS ROBUST-6G
Spyridon Georgoulas	NCSR Demokritos	Greece	SNS SAFE-6G
Harilaos Koumaras	NCSR Demokritos	Greece	SNS SAFE-6G, SNS UNITY-6G
Güneş Kesik	Ericsson Research	Turkey	SNS ROBUST-6G
Betül Güvenç Paltun	Ericsson Research	Turkey	SNS ROBUST-6G
Pietro G. Giardina	Nextworks	Italy	SNS ROBUST-6G
Gil Kedar	CTTC	Spain	SNS UNITY-6G
Selva Via Labrada	CTTC	Spain	SNS UNITY-6G

List of Reviewers

Editor Name	Institution	Country
Laurie Ibbs	Public Safety Communication Europe	Belgium
Sérgio Figueiredo	IPN - Instituto Pedro Nunes.	Portugal
Raul Orduna Urrutia	Vicomtech	Spain