Intent-Based Security: Bridging Natural Language Security Policies and Automated Network Defense

Pierre ALAIN[1], Ming SU[2], Daisy MUNSON[1], Issam IKHELEF[2], Fabien AUTREL[2], Guillaume DOYEN[2]
[1]Univ Rennes, IRISA, SOTERN [2]IMT Atlantique, IRISA, SOTERN

Identification of Use Cases

Modern military and critical infrastructure require security operators to define and maintain policies (typically through ISSPs) governing acceptable behaviors, access boundaries, and protection requirements. These policies are authored in natural language but translated manually into low-level configurations; any network evolution forces a new reassessment, a process that is error-prone, expert-dependent, and too slow for the pace of threats. Intent-Based Security (IBS) extends Intent-Based Networking (IBN) to the security domain, a field that introduces additional requirements around trust, deterministic enforcement, and resilience. It allows operators to express security objectives in natural language, delegating configuration derivation, deployment, and compliance verification to a system that autonomously verifies the complete enforcement chain.

For example, a tactical command post operates two segments: OPS (classified mission systems) and EXT (external partner link). The ISSP states: "Tactical systems must only communicate with authenticated external partners." At authoring time, IBS flags "authenticated" as ambiguous (the term could mean IP-based allow-listing or PKI-verified identity) and prompts the operator to clarify. In the absence of such a system, the translation may default to the weaker option, silently undermining the intended security guarantee. Separately, during a field exercise four months earlier, a temporary rule was added to allow a liaison officer's laptop direct access to OPS systems; it was never revoked. Three weeks into current operations, log analysis surfaces two overlapping patterns: the liaison laptop is establishing sessions to OPS at 3am and an OPS mission system is initiating connections to an EXT address recently reassigned to an unknown third party. IBS would surface these issues distinctly, classifying them into three structurally different mismatch types: a translation error, a missing intent, and two environment drift events, each presented with the appropriate urgency and a natural language explanation tied to the original policy text.

Primary users are security officers, network administrators, and Security Operations Center (SOC) analysts. The structural gap between natural language expression and device-level configuration is widened in defense contexts by tactical reconfigurations, coalition interoperability constraints, and degraded operations.

Proposed Technical Approach

We outline a research vision for an end-to-end IBS system bridging natural language expressions and automated network enforcement, from intent capture to policy assurance.
RFC 9315 [1] defines the intent lifecycle; RFC 9316 [2] proposes an intent classification methodology and reports the TM Forum autonomous network levels [3], from Level 0 (full manual control) to Level 5 (full autonomy). Leivadeas and Falkner [4] survey IBN architectures with limited security focus. Nile [5] and Lumi [6] provide a structured intent language and its NLP-based implementation in IBN systems. The design objective for IBS is Level 4 of the RFC 9316 taxonomy: a system where enforcement and monitoring are autonomous, but humans retain authority over policy definition and receive natural language notifications for decisions that fall outside defined bounds. Level 3 would require human approval for each enforcement action, incompatible with operational response speed, while Level 5 removes the accountability traceability required in defense contexts.

Formal security policy approaches (RBAC, ABAC, and similar access control frameworks) require manual translation from natural language norms to abstract representations, which does not scale to the volume and heterogeneity of access-control rules found in operational ISSPs or normative

technical standards. Cisco Security Cloud Control [Cisco 2026] automates intent-to-configuration deployment across multi-vendor firewalls, but requires structured input rather than natural language, and provides no formally verifiable correctness guarantees. In [7], Xu et al. propose security services via SDN/NFV instantiation without formal coverage of the derived configurations, and do not address monitoring or drift detection. Fogel et al. [8] present Batfish, a framework that verifies reachability, ACL correctness, and security properties across routing and firewall configurations; it operates on configurations as given and provides no mechanism for expressing, constructing, or monitoring network-wide security objectives. Tian et al. [9] propose Jinjing, which safely automates ACL configuration updates in large-scale WANs via a domain-specific update language; it operates at the level of ACL migration operations rather than high-level security objectives. Across these approaches, pre-IBN tools such as Batfish carry no notion of intent, while more recent systems operate at Levels 2 and 3: they automate specific operations but require structured technical input. IBS targets Level 4, covering the complete lifecycle from unstructured natural language norms to continuous intent assurance, with a deterministic, correct-by-construction enforcement chain and vendor-agnostic scope.

Several significant technical barriers must be addressed on this path:
(1) No security intent language exists; current approaches (whether IP-based ACLs, attribute-based policies, or group-based access rules) require operators to reason within predefined technical formalisms. None can be derived directly from natural language security documents such as ISSPs or 3GPP Technical Specifications, which present the additional challenge of cross-document anaphoric references. A language balancing expressivity, disambiguation and mappability to deterministic enforcement tools must be designed.
(2) Labeled corpora pairing natural language intents with formal representations are unavailable. The NEAT methodology [10], using open-weights models, bootstraps a corpus from a few expert examples; the same approach is being extended to normative documents (3GPP specifications), where selecting representative few-shot examples is complicated by the heterogeneity of access-control rules. Reliance on off-the-shelf LLMs introduces supply-chain trust concerns; mitigations include cross-model consistency validation and checking of generated pairs against manually validated references.
(3) Locally deployable open-weights models provide input confidentiality and output integrity guarantees that remote APIs cannot. Recent benchmarks show 32B-parameter models achieving approximately 85% of frontier performance on expert reasoning tasks [11], making local deployment viable for structured intent classification.
(4) With a well-defined intermediate intent representation, the downstream part of the chain is designed to be deterministic and correct by construction. Tools such as VEREFOO [12] synthesize firewall configurations that are guaranteed to implement the specified policy without approximation, in contrast to the probabilistic nature of the LLM-based upstream step. This property is essential in defense contexts: the system must not produce configurations that silently deviate from the intent.
(5) Continuous reconstruction of effective intent from configurations and logs must be compared against the expressed intent, and mismatches classified: a divergence may stem from an error in the translation step, from an environment change that invalidates a previously correct rule, or from an intent the operator never expressed; in the last case, the system must surface a correction proposal through the report stage of the cycle. ANIME [13] infers intents from forwarding behavior; Ribeiro et al. [14] extract security intents from firewall and NAT configurations. Both cover partial device views, reconstruct from configurations rather than from live traffic, and neither classifies mismatches between observed behaviour and declared intent, which remains an open problem.

Addressing these barriers constitutes the core research agenda on IBS, currently at an early stage with two PhD students and a postdoctoral researcher engaged on four work-streams: security intent language design, corpus construction adapting NEAT to the security domain, VEREFOO-based deterministic synthesis integration, and formalization of 3GPP access-control norms via few-shot translation with automated evaluation against a reference policy. Even partial progress would yield concrete operational benefits: IBS would surface within one log analysis cycle the silent divergences that currently persist undetected for weeks, classify each by root cause, and propose a corrective intent in natural language, closing the assurance loop without requiring operator expertise in low-level configurations.

References

[1] A. Clemm, L. Ciavaglia, L. Z. Granville, J. Tantsura, "Intent-Based Networking – Concepts and Definitions," RFC 9315, IRTF, October 2022.

[2] C. Li, O. Havel, A. Olariu, P. Martinez-Julia, J. Nobre, D. Lopez, "Intent Classification," RFC 9316, IRTF, October 2022.

[3] A. Boasman-Patel, D. Sun, Y. Wang, C. Maitre, J. Domingos, Y. Troullides, I. Mas, G. Traver, G. Lupo, "Autonomous Networks: Empowering Digital Transformation For The Telecoms Industry," TM Forum Whitepaper, Release 1.0, May 2019.

[4] A. Leivadeas and M. Falkner, "A Survey on Intent-Based Networking," IEEE Communications Surveys & Tutorials, vol. 25, no. 1, pp. 625–655, 2023.

[5] A. S. Jacobs, R. J. Pfitscher, R. A. Ferreira, L. Z. Granville, "Refining Network Intents for Self-Driving Networks," ACM SIGCOMM CCR, vol. 48, no. 5, pp. 55–63, 2018.

[6] A. S. Jacobs, R. J. Pfitscher, R. H. Ribeiro, "Hey, Lumi! Using Natural Language for Intent-Based Network Management," USENIX ATC, 2021.

[7] L. Xu, Y. Zhang, P. Chinprutthiwong, G. Gu, "Automatic Synthesis of Network Security Services: A First Step," IEEE ICCCN, 2023.

[8] A. Fogel, S. Fung, L. Pedrosa, M. Walraed-Sullivan, R. Govindan, R. Mahajan, T. Millstein, "A General Approach to Network Configuration Analysis," in Proc. 12th USENIX Symposium on Networked Systems Design and Implementation (NSDI), pp. 469–483, 2015.

[9] B. Tian, X. Zhang, E. Zhai, H. H. Liu, Q. Ye, C. Wang, X. Wu, Z. Ji, Y. Sang, M. Zhang, D. Yu, C. Tian, H. Zheng, B. Y. Zhao, "Safely and Automatically Updating In-Network ACL Configurations with Intent Language," in Proc. ACM SIGCOMM, pp. 214–226, 2019.

[10] D. Munson, P. Alain, G. Doyen, "NEAT: A Nile-English Aligned Translation Corpus Based on a Robust Methodology for Intent Based Networking," Computer Networks, vol. 271, 2025.

[11] DeepSeek-AI, "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning," arXiv:2501.12948, 2025.

[12] D. Bringhenti, G. Marchetto, R. Sisto, F. Valenza, J. Yusupov, "Automated Firewall Configuration in Virtual Networks," IEEE TDSC, vol. 20, no. 2, pp. 1559–1576, 2023.

[13] A. Kheradmand, "Automatic Inference of High-Level Network Intents by Mining Forwarding Patterns," ACM SOSR, 2020.

[14] R. H. Ribeiro, A. S. Jacobs, L. Zembruzki, R. Parizotto, E. J. Scheid, A. E. Schaeffer-Filho, L. Z. Granville, B. Stiller, "A Deterministic Approach for Extracting Network Security Intents," Computer Networks, vol. 214, 2022.

Industrial blog reference

[Cisco 2026] Cisco Security Blog, "Introducing Intent-Based Policy Management for Cisco Hybrid Mesh Firewall," January 2026.