2026 IEEE 50th Annual Computers, Software, and Applications Conference (COMPSAC) | 979-8-3315-4497-3/26/$31.00 ©2026 IEEE | DOI: 10.1109/COMPSAC69091.2026.00216

PANACEA: A Model-Based Framework for Self-Protecting Systems

Stefano Iannucci[‡], Emiliano Casalicchio[†], Francesco Guerra[*], Sara Pederzoli[*]
Matteo Paganelli[*], Tommaso Caiazzi[‡], Simone Albero[‡]
[‡]Roma Tre University, Italy; {stefano.iannucci, tommaso.caiazzi, simone.albero}@uniroma3.it
[†]Sapienza University of Rome, Italy; emiliano.casalicchio@uniroma1.it
[*]University of Modena and Reggio Emilia, Italy; {francesco.guerra, sara.pederzoli, matteo.paganelli}@unimore.it

***Abstract*—The increasing complexity of cyber threats and the growing autonomy of digital infrastructures demand security solutions that are adaptive, trustworthy, and resilient by design. This paper presents the outcomes of the PANACEA project, a PRIN 2022 initiative aimed at advancing the state of the art in model-based self-protecting systems. PANACEA integrates data-efficient intrusion detection, decision-theoretic intrusion response, and cyber-resilient autonomic management within a unified architectural framework. On the detection side, the project introduces adaptive models capable of operating under delayed and partially available labels, as well as a transformer-based semi-supervised learning approach that leverages self-supervised pretraining to significantly reduce reliance on labeled data while preserving competitive performance. On the response side, the project formalizes mitigation as a risk-aware decision process grounded in Timed Competitive Stochastic Games automatically synthesized from augmented Attack-Defense Trees (ADT), enabling cost-sensitive and explainable action selection. To address the vulnerability of the autonomic manager itself, PANACEA proposes a permissioned blockchain-based coordination layer that ensures consensus on system state and response decisions, tolerating the compromise of a bounded subset of components. All contributions are validated within a programmable cyber-range supporting controlled experimentation and end-to-end resilience assessment. The results demonstrate the feasibility of integrating adaptive learning, formal response planning, and distributed trust into a coherent self-protecting framework.**

***Index Terms*—Self-protecting systems, intrusion detection, intrusion response, model-based cybersecurity, decision-theoretic planning, adaptive anomaly detection, cyber-range experimentation, autonomous resilience.**

I. Introduction

PANACEA, short for *A Model-Based Framework for Self-Protecting Systems*, is a PRIN (Project of Relevant Interest for the Nation) 2022 research project funded by the Italian Ministry of University and Research (MUR)[1]. The project runs from September 28th, 2023 to February 28th, 2026, and investigates how adaptive intrusion detection, decision-theoretic intrusion response, and resilient autonomic management can be integrated into a coherent framework for self-protecting systems.

The project is coordinated by Stefano Iannucci from Roma Tre University, with Emiliano Casalicchio from Sapienza University of Rome and Francesco Guerra from the University of Modena and Reggio Emilia serving as co-principal investigators. The project team also includes Sara Pederzoli and Matteo Paganelli from the University of Modena and Reggio Emilia, and Tommaso Caiazzi and Simone Albero from Roma Tre University, who contributed to the machine-learning and system-prototyping activities.

The project outcomes include methodological contributions, prototype implementations, and publicly available software artifacts. The intrusion detection activities are documented through the NOCTOWL repository,[2] the NIDS framework repository,[3] and related PANACEA components.[4] The cyber-resilient autonomic management activities are supported by a blockchain-based coordination prototype.[5]

Overall, this paper summarizes the main objectives, technical results, dissemination outcomes, limitations, and future research directions of PANACEA. The contribution is organized as follows. Section II presents the objectives and expected outputs of the project. Section III discusses the relevance of the project to the COMPSAC symposia. Section IV reports the current status and main results. Section V discusses limitations and future work, and Section VI concludes the paper.

II. Objectives and Expected Outputs

The PANACEA project aims to advance the foundations of model-based self-protecting systems through the design and integration of data-efficient detection, decision-theoretic response, and resilient autonomic management within a unified framework. To this end, four main objectives are pursued:

- **Intrusion detection:** designing models capable of operating under realistic constraints such as delayed supervision, limited labeled data, and distribution shifts.
- **Intrusion response:** formalizing mitigation as a decision-theoretic process enabling risk-aware and cost-sensitive action selection.
- **Autonomic resilience:** ensuring the trustworthiness of the autonomic manager through distributed consensus mechanisms tolerant to partial compromise.
- **Experimental cyber-range infrastructure:** providing a programmable cyber-range for reproducible, end-to-

[1]https://prin.mur.gov.it/Pages/Index/119

[2]https://github.com/softlab-unimore/NOCTOWL
[3]https://github.com/Simone-Albero/nids-framework
[4]https://github.com/Marini97/PANACEA
[5]https://gitlab.com/uniroma3/compunet/networks/irs-blockchain

2836-3795/26/$31.00 ©2026 IEEE
DOI 10.1109/COMPSAC69091.2026.00216

Authorized licensed use limited to: UNIVERSITA MODENA. Downloaded on October 30,2026 at 10:09:56 UTC from IEEE Xplore. Restrictions apply.

end evaluation of detection, response, and coordination mechanisms.

A. Intrusion detection

Recent surveys on data-driven intrusion detection systems highlight the growing need for approaches that reduce reliance on labeled data while remaining effective under evolving traffic distributions [1], [2]. In line with this trend, the project has explored complementary approaches. On one side, adaptive tree-based models have been designed to operate under delayed and sampled label availability, integrating data shift detection mechanisms and sampling strategies to limit annotation costs in streaming settings [3]. On the other side, the project has introduced a hybrid semi-supervised approach based on transformer architectures, which leverages large volumes of unlabeled network traffic through self-supervised pretraining, using an autoencoding objective to learn domain-specific representations that are later refined with limited labeled data [4]. Together, these contributions substantiate the objective of data-efficient and practically deployable intrusion detection.

B. Intrusion response

Building on these detection capabilities, the project has also advanced automated intrusion response through decision-theoretic models. In particular, starting from prior work on model-based and near-optimal response planning [5]–[7], it has developed a methodology to synthesize Timed Competitive Stochastic Games from augmented Attack–Defense Trees (ADT), enabling a unified representation of attacker–defender interactions with temporal dynamics, uncertainty, and mitigation costs. Within this framework, intrusion response is formulated as a dynamic decision process, allowing the computation of risk-sensitive policies beyond static attack–response mappings [8]. This has led to response engines capable of selecting cost- and risk-aware countermeasures under operational constraints, while integrating detection outputs with structured system models to provide traceable and explainable mitigation strategies.

C. Autonomic resilience

A third objective tackled a long-standing challenge in autonomic computing: protecting the autonomic manager itself. In this context, several studies have emphasized the importance of safeguarding the manager [9]–[11] in order to ensure system-level trustworthiness. In line with this perspective, the project designed and prototyped a novel architecture for cyber-resilient, self-protecting systems, in which consensus mechanisms are leveraged to guarantee trustworthy detection and response decisions [12]. By leveraging a permissioned blockchain with Byzantine Fault Tolerant consensus, the architecture ensures agreement on both the system state derived from multiple intrusion detection instances and the response action selected among multiple intrusion response components. The resulting design tolerates the compromise of a bounded fraction of detection and response nodes and provides tamper-resistant logging of security-relevant events. Experimental evaluation of the prototype demonstrates that the overhead introduced by the blockchain-based coordination layer is negligible for typical event rates, while sustaining hundreds of security events per second even on modest virtualized infrastructures. This contribution transforms the abstract notion of cyber-resilience into a concrete architectural pattern supported by empirical evidence.

D. Experimental cyber-range infrastructure

A further objective was the realization of an experimental cyber-range infrastructure enabling reproducible, quantitative evaluation of self-protecting behaviors. The cyber-range supports controlled attack injection, workload variation, automated data collection, and performance logging, allowing systematic comparison of detection accuracy, label efficiency, response latency, mitigation effectiveness, and overall resilience. The infrastructure has been instrumental in validating both the detection models and the cyber-resilient autonomic manager, thereby providing an integrated experimental environment that connects algorithmic innovation with system-level validation.

III. Relevance

PANACEA spans multiple COMPSAC symposia, reflecting its interdisciplinary nature across cybersecurity, artificial intelligence, distributed systems, and autonomic computing.

At its core, PANACEA contributes to *Security, Privacy & Trust in Computing (SEPT)* by addressing the full lifecycle of self-protecting systems, from intrusion detection to response and resilient coordination. Adaptive models handling delayed and partially available labels, combined with decision-theoretic response planning, contribute to dependable security infrastructures. The blockchain-based consensus architecture further strengthens trust and fault tolerance by protecting the autonomic manager against partial compromise.

These capabilities are grounded in advances in *Applied Artificial Intelligence (AAI)* and *Data Sciences, Analytics & Technologies (DSAT)*. The project develops transformer-based semi-supervised learning pipelines that reduce reliance on labeled data through self-supervised pretraining. In parallel, adaptive tree-based models address streaming data, delayed supervision, and concept drift, bridging AI methodology and operational cybersecurity applications.

From a systems perspective, the project aligns with *Autonomous Systems (ASYS)* and *BlockChain (BEST)* by implementing a coherent monitor–analyze–plan–execute loop grounded in explicit system models and coordinated through permissioned consensus mechanisms. This design supports distributed decision-making with guarantees on integrity and consistency, even in partially compromised environments.

PANACEA further relates to *Networks, Communications, Internet & Web Technologies (NCIW)* through its focus on flow-based network analytics, and to *Software Engineering Technologies & Applications (SETA)* via the development of a programmable cyber-range for systematic experimentation. The availability of validated prototypes and scalability assessments also connects the project to *Industry* and *IT in Practice (ITiP)*, supporting technology transfer and practical adoption.

Authorized licensed use limited to: UNIVERSITA MODENA. Downloaded on October 30,2026 at 10:09:56 UTC from IEEE Xplore. Restrictions apply.

Overall, PANACEA provides a cross-cutting contribution that connects secure and trustworthy computing, applied AI, autonomous systems, blockchain coordination, and resilient software architectures within an experimentally validated cyber-defense framework.

IV. Current Status and Results

The PANACEA project has reached its final stage, with all major milestones achieved and the core architectural components implemented, and validated. At the time of writing, the project has transitioned from development to consolidation, with scientific dissemination and refinement of experimental artifacts representing the primary ongoing activities.

A. Technical Results

Intrusion detection. From the perspective of intrusion detection, the project has delivered both adaptive and data-efficient learning solutions. The semi-supervised transformer-based framework has been fully implemented and evaluated across multiple benchmark datasets, demonstrating that self-supervised pretraining on unlabeled flow data significantly reduces the number of labeled samples required to achieve competitive performance. Experimental campaigns have quantified this effect in terms of F1 score and labeled data savings, confirming that the hybrid approach achieves up to 98.8% of the performance of fully supervised models while using only half of the labeled data, and reaches comparable F1 scores with significantly fewer annotated samples [4]. In parallel, the adaptive tree-based model designed for delayed and sampled label availability has been validated in streaming settings under realistic conditions. The evaluation considers heterogeneous distribution shifts, including changes in feature distributions (covariate shift), class proportions (ratio shift), and the relationship between inputs and labels (concept shift). Under constrained supervision, characterized by a delay of 250,000 timestamps and a sampling rate of 20% of the labeled data, the model achieves an average F1 score of 70.6, outperforming state-of-the-art approaches and demonstrating robustness to delayed feedback and evolving data distributions [3].

Overall, these results confirm that the detection layer effectively meets the project's objectives in terms of adaptability, label efficiency, and operational feasibility in real-world intrusion detection scenarios.

Intrusion response. On the intrusion response side, the decision-theoretic response engine has been integrated with the detection modules through a structured knowledge layer encoding system state, attack semantics, and response cost models. Prototype implementations support risk-aware action selection and policy-driven mitigation strategies. Experimental evaluation within the cyber-range environment has demonstrated that response selection can be executed within bounded latency while preserving consistency with modeled constraints and operational policies.

Autonomic resilience. A major milestone achieved during the final project phase concerns the realization of a cyber-resilient autonomic manager based on permissioned blockchain technology. The architecture has been prototyped using a Byzantine Fault Tolerant consensus mechanism to coordinate multiple IDS and IRS instances. Experimental results indicate that the additional coordination overhead is limited and that the system can sustain hundreds of security events per second under realistic deployment assumptions. These findings validate the feasibility of protecting the autonomic manager itself, thus addressing a critical vulnerability in traditional self-protecting architectures.

Experimental cyber-range infrastructure. The cyber-range infrastructure has also reached operational maturity. It supports configurable network topologies, controlled attack injection, automated data collection, and repeatable benchmarking procedures. The platform has been used to conduct integrated experiments combining detection, response, and consensus-based coordination, thereby enabling end-to-end evaluation of resilience metrics. Reproducible experimental scripts, documented configurations, and publicly available code repositories ensure that the obtained results can be independently verified and extended.

B. Dissemination and Impact

In terms of dissemination and impact, the project has produced peer-reviewed publications covering adaptive intrusion detection under delayed labeling, semi-supervised transformer-based learning for label-efficient NIDS, and blockchain-based architectures for cyber-resilient self-protecting systems. Technologically, it has delivered functional prototypes of adaptive detection modules, a decision-theoretic response engine, and a consensus-based autonomic manager, together with an operational cyber-range for experimentation. Software artifacts, publicly available code repositories, and documented experimental protocols further contribute to reproducibility and technology transfer. Overall, the project has completed its planned research and development activities, delivering validated detection and response models, a resilient autonomic management architecture, and a programmable cyber-range for systematic experimentation. The current focus is on consolidating results, refining experimental analyses, and exploring follow-up research directions that build upon the established self-protecting framework.

V. Discussion, Limitations and Future Work

The PANACEA project demonstrates that adaptive learning, delay-aware supervision handling, formal response planning, and distributed trust mechanisms can be integrated into a coherent architecture for self-protecting systems. At the same time, the project highlights limitations and identifies concrete directions for future research.

Intrusion detection. Two complementary directions have been explored. First, semi-supervised transformer-based models improve data efficiency by leveraging unlabeled traffic through self-supervised pretraining, but their effectiveness depends on data representativeness, remains sensitive to domain shifts, and entails non-negligible computational costs. Future work will target adaptive pretraining criteria, lightweight or

Authorized licensed use limited to: UNIVERSITA MODENA. Downloaded on October 30,2026 at 10:09:56 UTC from IEEE Xplore. Restrictions apply.

distilled architectures, and increased robustness to adversarial settings. Second, adaptive tree-based models explicitly address streaming scenarios with delayed and partial labels, showing robustness to evolving distributions but requiring careful parameter tuning and raising scalability concerns. Future efforts will focus on controlling model growth over time, e.g., through data pruning and more stable adaptation strategies.

Intrusion response. The integrated decision-theoretic component enables risk-aware and cost-sensitive mitigation, supporting principled trade-offs and traceable decision-making. However, its effectiveness depends on the availability and accuracy of system and threat models, whose maintenance can be challenging in dynamic and heterogeneous environments. Scalability also remains a concern as the state and action spaces grow. Future work will focus on the adoption of scalable approximation techniques and automated extraction of system models from telemetry data.

Autonomic resilience. The integrated blockchain-based coordination layer enables consensus on detected events and selected countermeasures, ensuring consistent and trustworthy decisions even in the presence of partially compromised components. However, consensus protocols introduce serialization constraints and may limit scalability under high workloads. In addition, operational aspects such as secure configuration, key management, and governance policies remain critical to preserve system trustworthiness. Future work will investigate more scalable and adaptive coordination strategies, including parallel or sharded consensus mechanisms and formally verified smart contracts, to mitigate scalability constraints while balancing robustness and performance.

Experimental cyber-range infrastructure. The programmable cyber-range enables reproducible, end-to-end validation of detection, response, and coordination mechanisms, providing a controlled environment for attack injection, workload variation, and resilience assessment. However, such abstractions cannot fully capture the heterogeneity, legacy dependencies, and human operational factors of real-world infrastructures. Future work will focus on extending validation to larger-scale and more diverse environments, improving the assessment of external validity and practical deployment conditions.

VI. Conclusion

This paper presented the outcomes of the PANACEA project, a PRIN 2022 initiative advancing model-based self-protecting systems through the integration of adaptive intrusion detection, formal intrusion response, and cyber-resilient autonomic management. The results show that data-efficient learning approaches, combining semi-supervised transformer models and delay-aware adaptive trees, can operate effectively under realistic constraints such as limited and delayed supervision. In parallel, modeling intrusion response as a decision-theoretic process and introducing a permissioned blockchain-based coordination layer strengthen both the rationality and trustworthiness of autonomous defense.

Experimental validation in a programmable cyber-range demonstrates the feasibility of unifying learning, planning, and distributed consensus within a coherent architecture. While challenges related to scalability, long-term adaptability, and deployment complexity remain, PANACEA provides a solid foundation for the design of next-generation cyber-resilient self-protecting systems, supporting a vision of autonomous cyber defense grounded in both formal methods and data-driven intelligence.

References

[1] D. Chou and M. Jiang, "A survey on data-driven network intrusion detection," *ACM Comput. Surv.*, vol. 54, no. 9, pp. 182:1–182:36, 2022.

[2] R. Chinnasamy, M. Subramanian, S. V. Easwaramoorthy, and J. Cho, "Deep learning-driven methods for network-based intrusion detection systems: A systematic review," *ICT Express*, vol. 11, no. 1, pp. 181–215, 2025.

[3] S. Pederzoli, M. Paganelli, M. L. Contalbo, R. Benassi, D. Tiano, S. Iannucci, and F. Guerra, "Noctowl: Adaptive tree-based model for network anomaly detection under delayed and sampled label availability," *IEEE Access*, vol. 13, pp. 197 899–197 911, 2025.

[4] S. Albero, T. Caiazzi, S. Iannucci, P. Merialdo, and R. Torlone, "Leveraging semi-supervised learning to reduce labeled data requirements in intrusion detection," in *2025 IEEE 49th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, 2025, pp. 262–267.

[5] S. A. Zonouz, H. Khurana, W. H. Sanders, and T. M. Yardley, "RRE: A game-theoretic intrusion response and recovery engine," *IEEE Transactions on Parallel and Distributed Systems*, 2014.

[6] X. An, F. Lin, S. Xu, L. Miao, and C. Gong, "A novel differential game model-based intrusion response strategy in fog computing," *Security and Communication Networks*, 2018.

[7] S. Iannucci and S. Abdelwahed, "Model-based response planning strategies for autonomic intrusion protection," *ACM Transactions on Autonomous and Adaptive Systems*, 2018.

[8] T. Caiazzi, S. Iannucci, V. Marini, M. Foschi, and R. Torlone, "From attack trees to timed stochastic games: a novel intrusion response approach," *Computers & Security*, p. 104834, 2026.

[9] D. M. Chess, C. C. Palmer, and S. R. White, "Security in an autonomic computing environment," *IBM Systems journal*, vol. 42, no. 1, pp. 107–118, 2003.

[10] E. Yuan, N. Esfahani, and S. Malek, "A systematic survey of self-protecting software systems," *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, vol. 8, no. 4, pp. 1–41, 2014.

[11] F. Flammini, C. Alcaraz, E. Bellini, S. Marrone, J. Lopez, and A. Bondavalli, "Towards trustworthy autonomous systems: Taxonomies and future perspectives," *IEEE Transactions on Emerging Topics in Computing*, 2022.

[12] T. Caiazzi, S. Iannucci, V. Marini, D. Pennino, M. Pizzonia, and R. Torlone, "A novel architecture for cyber-resilient self-protecting systems based on blockchain," in *2025 IEEE 49th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, 2025, pp. 1–8.

Authorized licensed use limited to: UNIVERSITA MODENA. Downloaded on October 30,2026 at 10:09:56 UTC from IEEE Xplore. Restrictions apply.