Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Position Paper:

Active cybersecurity: vision, model, and key technologies*◇

Xiaosong ZHANG†‡, Yukun ZHU†, Xiong LI†, Yongzhao ZHANG, Weina NIU, Fenghua XU, Junpeng HE, Ran YAN, Shiping HUANG†

School of Computer Science and Engineering, University of Electronic Science and Technology of China, Sichuan 611731, China

†E-mail: johnsonzxs@uestc.edu.cn; maln3bul2@gmail.com; lixiong@uestc.edu.cn; shiping@std.uestc.edu.cn

Received Jan. 23, 2025; Revision accepted Apr. 22, 2025; Crosschecked June 9, 2025; Published online June 27, 2025

Abstract: Noncooperative computer systems and network confrontation present a core challenge in cyberspace security. Traditional cybersecurity technologies predominantly rely on passive response mechanisms, which exhibit significant limitations when addressing real-world complex and unknown threats. This paper introduces the concept of "active cybersecurity," aiming to enhance network security not only through technical measures but also by leveraging strategy-level defenses. The core assumption of this concept is that attackers and defenders, in the context of network confrontations, act as rational decision-makers seeking to maximize their respective objectives. Building on this observation, this paper integrates game theory to analyze the interdependent relationships between attackers and defenders, thereby optimizing their strategies. Guided by this foundational idea, we propose an active cybersecurity model involving intelligent threat sensing, in-depth behavior analysis, comprehensive path profiling, and dynamic countermeasures, termed SAPC, designed to foster an integrated defense capability encompassing threat perception, analysis, tracing, and response. At its core, SAPC incorporates theoretical analyses of adversarial behavior and the optimization of corresponding strategies informed by game theory. By profiling adversaries and modeling confrontation as a "game," the model establishes a comprehensive framework that provides both theoretical insights into and practical guidance for cybersecurity. The proposed active cybersecurity model marks a transformative shift from passive defense to proactive perception and confrontation. It facilitates the evolution of cybersecurity technologies toward a new paradigm characterized by active prediction, prevention, and strategic guidance.

Key words: Active cybersecurity; Intelligent threat sensing; In-depth behavior analysis; Comprehensive path profiling; Dynamic countermeasures

https://doi.org/10.1631/FITEE.2500053 **CLC number:** TP393.08

1 Introduction

The continuous advancement of global digital technologies has transformed cyberspace into a critical domain for national competition. As cyberspace becomes more important in politics, economics, and military matters, cybersecurity has become essential for protecting national interests and keeping society stable. In recent years, cyberattacks have become not only more frequent but also increasingly sophisticated. The advent of artificial intelligence (AI) and big data technologies (Kaur et al., 2023; Rajapaksha et al., 2023) has enabled the attacker to apply highly customized attacks that are more difficult to detect and are executed with greater speed, thus impacting a broader range of system (Han et al., 2021).

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. U2336204 and 62372086) and the Natural Science Foundation of Sichuan Province, China (Nos. 2024NSFSC0004 and 2025ZNSFSC0500)

◇ Special Topic: The 27th Annual Meeting of the China Association for Science and Technology

ORCID: Xiaosong ZHANG, https://orcid.org/0000-0001-9886-1412

In response to the increased cybersecurity threats, various cybersecurity models have been proposed, providing not only theoretical foundation but also practical guidance for cybersecurity systems to counter evolving threats. These models are mainly defense-centric ones, focusing on enhancing the defensive capabilities (e.g., the protection, detection, and response (PDR) model (Schwartau, 1998), P2DR (Li DP et al., 2014), and PDR + recovery PDRR (Yang Y et al., 2024)) of cybersecurity systems in dynamic environments. However, these models focus mainly on how systems can defend themselves internally. They lack in-depth insights into the behaviors, motivations, and strategies of attackers. This places the defender in a passive position during attack–defense confrontations.

To break away from the traditional "defense-centric" mindset, several models centered on attack–defense confrontation have been developed. These models examine attacks from the perspective of the attacker, enabling the defender to gain a deeper understanding of the attack mechanisms. The "find, fix, track, target, engage, and assess" (F2T2EA) framework (Tirpak, 2000) shows how to quickly spot threats, stop them from spreading, and respond effectively. Building on this, the "seven-step kill chain" model (Sun S et al., 2023) helps the defender by breaking down the key stages in which an attack operates. The adversarial tactics, techniques, and common knowledge (ATT&CK) framework (Strom et al., 2020), created by the US-based Mitre Corporation, provides a detailed map of attack methods and behaviors. This framework has become a popular tool for analyzing complex threats and testing defense systems. However, these models lack in-depth analysis of the strategies applied in dynamic attack–defense confrontations.

The attack–defense confrontation between the attacker and the defender resembles a dynamic "game," where both sides continuously adapt their strategies in response to the actions of the opposing side. The attacker continuously adapts the attack strategies in response to the vulnerabilities discovered in systems and the gaps identified in defenses (Crandall et al., 2005). For instance, the attacker may apply multistage attacks to infiltrate systems undetected or rapidly switch tactics when the defender updates the security measures. In turn, defenders must continuously enhance their strategies to prevent evolving threats. For example, the defender may utilize upgrading detection algorithms, deploy threat awareness systems, or enhance dynamic response mechanisms. Therefore, the techniques on both sides are constantly evolving, leading to a dynamic and adversarial game in the cybersecurity field. The core of this game lies in how both sides use information, technology, and timing. The attacker seeks to maximize disruption or infiltration gains, while the defender aims to minimize attack success and impact before threats emerge. This dynamic adversarial relationship introduces uncertainty and complexity into cybersecurity, requiring systems to be both flexible and predictive.

Given the dynamic and adversarial nature of cyberspace attack–defense confrontations, several limitations of cybersecurity models remain as follows:

1. Unknown threat imperception. Traditional cybersecurity models primarily rely on passive defense strategies, using predefined rules and pattern matching to address known threats. However, this approach often struggles to detect new or sophisticated attacks promptly, resulting in delayed responses. Although models centered on attack confrontation offer perspective insights of the attacker, providing the defender with some initiative, they still lack a holistic view. Consequently, these models are less effective in predicting attacks, formulating countermeasures, and mitigating threats, particularly as attacks become more complex, concealed, and rapidly evolving.

2. Limited strategic coordination. Current cybersecurity models primarily emphasize advanced technical methods, such as attack detection and pattern matching. However, they frequently overlook the integration of these tools within broader defense strategies. This disconnection between technical tools and strategic planning restricts defense effectiveness, particularly in the face of evolving threats. Consequently, ensuring sustained and comprehensive security becomes increasingly challenging.

3. Lack of continuous optimization. Current cybersecurity models emphasize detecting and responding to immediate attacks but generally neglect long-term threats and system enhancements. This limitation hampers their capacity to adapt to emerging threats and sustain effective security over extended periods.

To address the aforementioned limitations, this

paper proposes an active cybersecurity model centered on four core elements: intelligent threat sensing, in-depth behavior analysis, comprehensive path profiling, and dynamic countermeasures (SAPC). Under the assumption that both the attacker and the defender are rational decision-makers who aim to maximize their respective objectives, we incorporate game theory (Tan et al., 2023) as an analytical approach, thus enabling the SAPC model to conceptualize cyber space attack–defense confrontation as a continuously evolving strategy optimization process.

Besides preventing known threats, the SAPC model focuses on constructing a defense system capable of proactively countering attacks by analyzing and predicting attacker behavior. This model establishes a comprehensive cycle, encompassing threat detection, prevention, behavior analysis, path profiling, and counteraction with the application of game theory. This application of game theory continuously optimizes defense strategies, allowing the defender to develop a more profound understanding of the attacker and sustain strategic dominance.

The SAPC model provides innovative theoretical perspectives and practical approaches for addressing contemporary cyber threats. By integrating its four core elements into a continuous cycle, it grants the defense strategies with more flexibility and adaptability in complex scenarios, thereby strengthening the overall security. In reality, it can contribute to national security by safeguarding critical infrastructure, and it supports businesses by mitigating risks during digital transformation.

The main contributions of this work are as follows:

1. We integrate the idea of game theory into the field of cybersecurity, under the assumption that both attackers and defenders are rational decision-makers aiming to maximize their respective objectives.

2. We propose a forward-looking vision for active cybersecurity. This vision focuses on improving defense strategies and understanding attack patterns to build a modern active security framework.

3. We propose SAPC, an active cybersecurity model with four key elements: intelligent threat sensing, in-depth behavior analysis, comprehensive path profiling, and dynamic countermeasures. These elements work together in a continuous cycle to detect and contain threats.

4. We analyze attack–defense confrontation through dynamic game theory, utilizing local and global games to optimize active security strategies, thus advancing both theoretical frameworks and practical applications in modern cybersecurity.

The explanation of core terms in this paper is given in Appendix.

2 Related works

The evolution of cybersecurity models is a gradual process, driven by the increasing complexity of network attacks and the evolving nature of threat landscapes. These models can be grouped into two main categories: defense enhancement models and attack confrontation models.

2.1 Defense enhancement models

Defense enhancement models primarily focus on strengthening passive defense mechanisms to improve threat detection, prevention, and mitigation. These models incorporate technologies such as intrusion detection systems (IDSs), firewalls, and encryption to enhance the security infrastructure. A key characteristic of these models is the continuous monitoring and the adaptation of defense strategies to address evolving threats. Furthermore, defense enhancement models emphasize a layered security approach, deploying multiple protective measures throughout the network to enhance resilience and reduce vulnerabilities to attacks.

1. PDR model (Schwartau, 1998). Introduced by the US-based company intelligent security systems (ISSs), the PDR model demonstrates a structured approach to cybersecurity defense through its three core components: protection, detection, and response. However, the model has several notable limitations, including the lack of systematic security policy guidance, weak coordination among its components, and challenges in dynamically adapting defense strategies to real-time security contexts.

2. P2DR model (Li DP et al., 2014). The P2DR model was introduced to enhance the flexibility and adaptability of the PDR model. It incorporates a comprehensive array of protection tools and integrates overarching security strategy guidance, allowing for dynamic optimization of the defense system through adaptive detection and response mechanisms. However, the rapid proliferation of cloud

computing and mobile Internet has resulted in increasingly dispersed and complex attack methods. Consequently, the P2DR model faces challenges such as slow response and limited business continuity recovery capabilities in distributed environments.

3. PDRR model (Yang Y et al., 2024). The PDRR model introduces a "recovery" component to the original PDR framework, ensuring rapid business continuity and comprehensive restoration following an attack. While the PDRR model enhances post-event recovery capabilities by expanding its functions, its weaknesses in preventing threat identification lead to the evolution of the PDRR + audit analysis (PDR2A) model.

4. PDR2A model (Gao, 2012). The PDR2A model enhances vulnerability identification and prevents defense capabilities through an audit analysis module. However, as an event-driven framework, it lacks dynamic real-time threat management, which leads to the development of the next-generation models.

5. IPDRR model (Zhang X et al., 2023). Building on the PDR2A model, a complete security assurance framework named IPDRR was introduced to cover the entire lifecycle. This framework integrates five core capabilities, risk identification, security defense, detection, response, and recovery, significantly enhancing threat management and real-time response across the lifecycle.

6. APPDRR model (Xu XZ et al., 2024). The APPDRR model further incorporates a dynamic security policy module, creating a closed-loop protection system that spans from risk assessment to real-time response.

7. Intrinsic security (Sabnis et al., 2012). Intrinsic security focuses on developing active defense mechanisms within the system itself to enhance its resilience. In contrast to the passive detection applied by traditional models, it offers a more active approach to threat detection. Intrinsic security is a security concept aimed at strengthening self-defense capabilities, sharing a similar philosophy with mimetic security.

8. Mimetic security (Wu, 2016). Mimetic security is a concept rooted in "diversity" and "dynamism." Inspired by the mimetic characteristics of biological systems, it increases the cost of attacks through dynamic changes and randomization mechanisms. The introduction of endogenous and mimetic security offers strategic guidance for the development of cybersecurity models, promoting a shift from passive defense to active adaptation. However, both intrinsic and mimetic security models face limitations in today's highly adversarial network environment. The active detection and decision-making processes in endogenous security can result in high resource consumption, thereby reducing the overall defensive efficiency. Meanwhile, mimetic security struggles to effectively constrain the offensive behaviors of attackers. Fundamentally, both approaches focus heavily on internal defense and fail to provide comprehensive insights into attackers' behaviors, motivations, and strategies, which thus limits their ability to take the initiative in complex attack–defense confrontations.

9. LMSanitator framework (Wei et al., 2024). The LMSanitator framework exemplifies defense-enhanced cybersecurity models deployed in concrete application scenarios. As a representative implementation under the defense enhancement model category, this approach applies runtime behavioral verification mechanisms to neutralize malicious prompts while maintaining baseline model functionality. However, it inherits fundamental limitations common to enhancement-oriented methodologies: (1) insufficient integration of adversarial perspective analysis, specifically lacking systematic modeling of attack decision-making processes; (2) contextual inflexibility, which restricts its adaptability to evolving attack surfaces beyond predefined threat models.

2.2 Attack confrontation models

Compared to traditional defense enhancement models, the active cybersecurity framework, centered on attack–defense confrontation, represents the forefront of cybersecurity research. The core of the active cybersecurity concept lies in transcending the traditional "defense-centered" approach, emphasizing a dynamic balance between defense and offense. It integrates attack behaviors, strategies, and intentions into the design of security systems, thereby creating a more comprehensive framework.

In recent years, cybersecurity models inspired by this concept have emerged, providing new technical support and theoretical foundations for modern cybersecurity systems. These models analyze attacks from the attackers' perspective, revealing key stages of the attack chain and offering

comprehensive threat intelligence to defenders.

1. The F2T2EA model (Tirpak, 2000). The F2T2EA attack architecture, introduced in 1996, focuses on the rapid discovery of targets, threat locking, precise strikes, and effect evaluation. Its clear process and efficient response capabilities laid the foundation for subsequent attack–defense confrontation models.

2. Kill chain model (Sun S et al., 2023). The "seven-step kill chain" model, proposed in 2011, emphasizes the importance of identifying and analyzing key steps in the attack chain to develop effective defense strategies.

3. ATT&CK model (Strom et al., 2020). In 2013, the ATT&CK model, introduced by the US-based Mitre Corporation, systematized attack paths and behaviors through a matrix-based tactical and technical knowledge base. This model is widely used to assess attack–defense capabilities, analyze advanced persistent threats (APTs), and conduct threat hunting. This model helps defenders gain deeper insights into attack strategies and behaviors while enhancing the intelligence and proactivity of defense systems.

4. WPDRRC model (Yao, 2010). The WPDRRC model, which represents the attack–defense confrontation approach, integrates personnel, strategy, and technology. It proactively alerts defenders to potential threats and weakens the attacker's operational capabilities through counterattack mechanisms, overcoming the limitations of traditional defense strategies.

5. CARTA model (Jiang X, 2020). The CARTA architecture enables real-time responses to network changes through dynamic risk assessments and continuous adjustments, significantly enhancing the flexibility of the defense system.

6. Shield model (Fowler et al., 2020). The Shield model actively disrupts attackers' actions and decision-making efficiency through deception technologies and induction strategies.

7. MTD model (Cai et al., 2016). This model dynamically changes the system and network states across multiple dimensions, including time, space, and the physical environment, to increase the difficulty for attackers to target and exploit vulnerabilities. However, it features high resource consumption and lack of coordination among components.

8. SARPPR model (Fang et al., 2024). The SARPPR model focuses on full lifecycle defense, combining the "guard, self-defense, and iteration" modes to optimize defense mechanisms and enhance the perception and deception of potential attackers.

9. Space Odyssey framework (Willbold et al., 2023). The Space Odyssey framework serves as a representative implementation of attack confrontation models within the emerging domain of space system security. By extending the core principle of adaptive attack surface modeling to satellite command chains, this framework pioneers a systematic methodology for countering heterogeneous attack vectors in orbital environments. The framework demonstrates the technical feasibility of applying attack confrontation models to space security scenarios while exposing systemic deficiencies in current space cybersecurity research across dimensions such as cross-domain coordination and real-time response.

2.3 Limitations of existing works

Despite advancements in dynamic defense, real-time response, and attacker behavior analysis, current cybersecurity models continue to exhibit significant limitations. Table 1 provides a summary of the advantages and disadvantages of the models discussed, with additional evaluations across four key dimensions: proactivity, flexibility, comprehensiveness, and lightness. Proactivity refers to the consideration of the attacker's perspective, while flexibility denotes the model's ability to adapt to dynamically changing environments. Comprehensiveness reflects the breadth of defense dimensions and stages considered. Lightness pertains to resource consumption at any given time, with higher lightness corresponding to lower resource consumption and vice versa.

As shown in Table 1, defense enhancement models exhibit relatively low proactivity but generally have higher lightness compared to attack confrontation models. This suggests that current defense enhancement models lack a thorough understanding of attacker behaviors, motivations, and strategies, thereby limiting their ability to take a proactive role in attack–defense scenarios. On the other hand, the attack–defense confrontation model requires proactive detection and engagement in adversarial games with the attacker, which leads to greater resource consumption. Furthermore, as the models evolve, both flexibility and comprehensiveness

Table 1 Summary of each cybersecurity model

Category	Model	Advantage	Disadvantage	P	F	C	L
Defense enhancement	PDR (Schwartau, 1998)	Structured defense	Lack of dynamic adjustment	○	○	○	●
	P2DR (Li DP et al., 2014)	Security strategy for dynamic optimization	Slow response in distributed environments	○	○	○	●
	PDRR (Yang Y et al., 2024)	Recovery phase for full restoration	Weak pre-attack threat identification	○	○	○	●
	PDR2A (Gao, 2012)	Auditing vulnerability	Limited real-time threat management	○	◐	◐	●
	IPDRR (Zhang X et al., 2023)	Lifecycle framework	High complexity and cost	○	◐	◐	◐
	APPDRR (Xu XZ et al., 2024)	Closed-loop protection	Limited dynamic adaptability	○	◐	●	◐
	Intrinsic security (Sabnis et al., 2012)	Active threat awareness	High resource consumption	◐	●	●	○
	Mimicry security (Wu, 2016)	Increased attack difficulty	Focusing on internal defense	○	●	◐	◐
	LMSaniator (Wei et al., 2024)	Runtime behavior verifaction	Insufficient adversarial perspective and contextual inflexibility	◐	◐	◐	●
Attack confrontation	F2T2EA (Tirpak, 2000)	Structured attack process	Lack of dynamic adaptability	◐	○	◐	●
	Kill chain (Sun S et al., 2023)	Modeling attack	Limited adaptability	◐	◐	◐	◐
	ATT&CK (Strom et al., 2020)	Matrix attacker behavior	Struggling with unknown threats	●	◐	●	◐
	WPDRRC (Yao, 2010)	Active warning and counterattack	Risk of greater retaliation	●	◐	●	○
	CARTA (Jiang X, 2020)	Real-time response	High resource consumption	●	●	◐	○
	Shield (Fowler et al., 2020)	Deception attacker	Limited proactive feature	◐	◐	◐	◐
	MTD (Cai et al., 2016)	Dynamic network and system	High resource consumption and low coordination	●	◐	●	○
	SARPPR (Fang et al., 2024)	Full lifecycle protection	High resource consumption	●	●	●	○
	Space Odyssey framework (Willbold et al., 2023)	Adaptive attack surface modeling	Cross-domain coordination and real-time reponse	●	●	◐	◐
Proposed model	SAPC	Full-stage dynamic game	–	●	●	●	●

P represents proactivity, F represents flexibility, C represents comprehensiveness, and L represents lightness. ○ represents low, ◐ represents medium, and ● represents high

are expected to increase, though this comes at the expense of lightness.

In conclusion, existing models fail to achieve an optimal balance among proactivity, flexibility, comprehensiveness, and lightness. Moreover, many models rely on static or single-round game strategies, which are inadequate for supporting continuous dynamic optimization. Consequently, they are ill-equipped to address the increasingly complex, covert, and rapidly evolving threats in today's highly dynamic network environments. Therefore, research should focus on developing a comprehensive lifecycle dynamic game model that enhances the ability to predict attacker intentions and dynamically adjust defense systems, thereby enabling comprehensive optimization and systematic progress in attack–defense confrontations. Our model, namely, SAPC, is designed based on this motivation, and we provide a more detailed explanation in the following sections.

3 Active cybersecurity vision

Considering the limitations of the current active cybersecurity model in addressing dynamic games and comprehensive defense mechanisms, this paper proposes a forward-looking perspective on active cybersecurity. This perspective emphasizes not only the refinement of defense strategies but also an in-depth examination of malicious behavior patterns and technical attributes, aiming to develop a

contemporary cybersecurity framework grounded in dynamic game theory.

3.1 Concepts and definitions

Active cybersecurity seeks to build a systematic, active defense capability, encompassing a full spectrum of functions from threat detection and analysis to localization and mitigation (Shi et al., 2024). By combining active and passive security strategies, it aims to predict and preempt potential risks before threats materialize (Hu et al., 2024). This approach implements measures to minimize potential losses across economic, political, and military domains, ensuring comprehensive cyberspace protection with respect to its confidentiality, integrity, availability, and provability.

Active cybersecurity can be characterized as a systematic defense framework that combines active and passive strategies and technologies, specifically designed to counteract network threats and attacks effectively (Hand et al., 2013). This framework strengthens the ability to address sudden network incidents by proactively detecting and forecasting potential threats. Its fundamental components include the following:

1. Real-time threat perception: utilizing multiple detection rules to identify potential threats in infrastructures and networks.

2. In-depth data analysis: investigating evolving malicious behaviors and patterns to uncover underlying trends.

3. Threat capability assessment: analyzing the capabilities of the attacker and evaluating the potential impact and destructiveness of the attack actions.

4. Holistic protection: deploying available resources to formulate and implement strategies for threat tracking, tracing, and mitigation, thereby establishing a resilient and efficient cybersecurity framework. Compared with the MTD model, which possesses defects in resource consumption and adaptability, active cybersecurity presents a holistic protection through collaboration and predictive strategies.

3.2 Objectives of active cybersecurity

The objectives of active cybersecurity encompass four fundamental dimensions: manageable security, traceable security, provable traceability, and controllable disposal. These dimensions align with the holistic management of network assets, the audit and analysis of complex network behaviors, the reliable traceability of attack activities, and the accurate mitigation of targeted threats, separately.

1. Manageable security. This involves comprehensively managing network assets, potential threats, and associated risks to build an integrated protection system focused on proactive prediction and precise response. This facilitates early warning mechanisms and targeted protective measures to reduce the likelihood and impact of security incidents, ensuring the safety and stability of system operations in complex network environments. Furthermore, manageable security incorporates advanced threat detection capabilities based on dynamic game theory, enabling continuous prediction and analysis of newly identified attack vectors by adversaries. Insights drawn from the overall security posture are leveraged to preemptively counter emerging threats.

2. Traceable security. This involves systematically recording, tracking, and auditing the entire behavioral chain of complex network activities and cross-domain covert operations, thereby improving the visibility and transparency of network actions. Traceable security ensures that all activities within the network are accurately and comprehensively documented and analyzed, providing strong data support for prediction, decision-making, and response to potential security incidents, while enhancing the efficiency of network security operations. It develops advanced capabilities for analyzing hidden attacks based on dynamic game theory, enabling the ongoing identification of various attack methods designed to evade detection by integrating and analyzing multidimensional data, thereby improving the detection of highly concealed threats.

3. Provable traceability. This involves utilizing information obtained through traceable security to reconstruct the attack action path, including the attack origin, propagation route, exploited vulnerabilities, and final target. This process generates credible evidence for comprehensive attack chain tracking, providing robust technical and legal support for evidence collection, accountability, and litigation. It ensures transparency, fairness, and traceability throughout the threat response process (Jiang JG et al., 2018). Provable traceability further develops precise tracing and defense capabilities based on

dynamic game theory, enabling continuous tracking of attack paths, reconstruction of attack chains, and detailed analysis of attacker behavior trajectories. These measures enhance the accuracy of source tracing and guide defenders in deploying more targeted protection strategies against covert attacks.

4. Controllable disposal. This feature involves developing comprehensive control capabilities across all stages of threat detection, response, containment, and recovery. Specifically, controllable disposal applies precise countermeasures against threat behaviors, ensuring swift system recovery after attack while minimizing the impact of business disruptions. Additionally, it incorporates dynamic game-based defense strategy adjustment capabilities to counter attackers, adaptively refining and optimizing defense and countermeasure strategies based on response effectiveness. This approach continually increases the cost and complexity for attackers, achieving the objective of "enhancing system resilience through attack–defense confrontation."

3.3 Significance and strengths

Active cybersecurity integrates manageable security, traceable security, provable traceability, and controllable disposal, advancing network security through both technical measures and strategic defenses. This subsection explores its significance and key strengths.

3.3.1 Significance

The significance of active cybersecurity lies in establishing a next-generation immune capability for defense mechanisms through preemptive threat perception, in-depth source analysis, dynamic strategy reconstruction, and immunity achievement.

1. Preemptive threat perception: By utilizing AI and attack path prediction to monitor network traffic, user behavior, and system vulnerabilities (Sun C et al., 2022), potential attack actions and security events are anticipated, enabling a transition from passive defense to proactive warning for rapid response.

2. In-depth source analysis: Through the examination of attack behaviors and vulnerability characteristics, in-depth source analysis enables the identification of threat origins and propagation paths. Subsequently, attack chains, patterns, and strategies are analyzed to inform defense decision-making processes, facilitating the development of risk control measures and strengthening response capabilities.

3. Dynamic strategy reconstruction: Security policies and network configurations are continuously adjusted in response to evolving threats, thereby enhancing the adaptability of detection systems and ensuring the relevance of defensive measures to promptly identify and mitigate complex intrusions.

4. Immunity achievement: By leveraging dynamic game-based confrontation, flexible countermeasures tailored to specific attack types are developed by analyzing the objectives and strategies of the attacker. Consequently, defense mechanisms are strengthened, leading to the gradual establishment of an adaptive and self-recovering security immune system.

3.3.2 Strengths

The strengths of active cybersecurity are reflected in the following five aspects:

1. Continuous confrontation: From the perspective of the defender, four essential capabilities are developed: threat perception, attack analysis, precise tracing and protection, and defense strategy adaptation. These capabilities bolster the security resilience and stress tolerance of the system, thereby establishing a sustained advantage in addressing persistent threats within complex network environments.

2. Forward defense: By utilizing threat prediction and prevention, real-time dynamic perception, and the optimization of protection mechanisms, potential threats to critical assets are identified proactively. Attack paths are monitored in real time, and defense strategies are dynamically adjusted according to the characteristics of the threats. This approach shortens the detection and resolution time window, minimizes potential losses, and improves the overall effectiveness of cybersecurity.

3. Full-link controllability: Comprehensive protection is implemented throughout the threat lifecycle, from detection to countermeasures, including prior awareness, intelligent analysis, data tracking, forensic tracing, and attack mitigation. This ensures a coordinated and resilient defense mechanism.

4. Systematization and collaboration: By integrating multilayer security modules and sharing threat intelligence, cross-domain security gaps are eliminated, thereby enhancing overall protective capabilities.

5. Intelligence and standardization: On one hand, intelligent and automated technologies reduce human involvement, optimizing the security operations and the decision-making; on the other hand, standardized threat detection processes improve detection efficiency, minimize false positives and missed detections, and ensure precise responses through a unified resource allocation mechanism.

4 Proposed SAPC model

As cyberattacks become increasingly sophisticated, traditional defense mechanisms encounter significant challenges in maintaining effectiveness. This section introduces an active cybersecurity model, namely, SAPC, based on game theory, which dynamically refines defense strategies through four elements, namely, intelligent threat sensing, in-depth behavior analysis, comprehensive path profiling, and dynamic countermeasures, thereby addressing the evolving nature of network security threats. The theoretical foundations and key components of these elements are introduced herein, offering novel perspectives and methodologies for enhancing cybersecurity defense.

4.1 Overview

In recent years, the landscape of cyberattacks and defenses has significantly evolved compared to the previous decade. Cyberattacks have grown increasingly sophisticated and covert in nature. The focus of these attacks has shifted from targeting personal privacy to critical national infrastructure. Cyberspace has become one of the primary arenas for strategic competition among major powers. In the contemporary international landscape, various nations actively participate in activities such as executing surveillance operations, infiltrating digital networks, and initiating cyberattacks. This is evident particularly in the emergence of advanced threats such as APTs, ransomware, and supply chain attacks. Consequently, the defense sector faces significant challenges, necessitating the continuous evolution of defensive strategies to counter increasingly sophisticated attack methodologies.

The key challenges in cybersecurity encompass information asymmetry, the variability of attack strategies, and limitations in defensive resources. Attackers typically hold an informational advantage and can adapt their tactics flexibly through covert strategies, while defenders must close this gap through continuous learning and intelligence gathering. Cyberattacks exhibit considerable volatility, as attackers can rapidly modify their strategies in response to defensive measures. The existence of multiple attack vectors and uncertainty regarding the number of attacking agents further complicates the defensive challenge. The strategic dynamics of this offensive–defensive interaction requires defenders to respond adaptively to optimize the overall defensive effectiveness.

Game theory provides a systematic model for cybersecurity, facilitating the optimization of strategies for both attackers and defenders under the constraints of limited time and resources. Network security attacks and defenses form a dynamic, evolving game-theoretic process. Attackers and defenders compete in resource allocation and strategy selection, each aiming to maximize their own benefits. In the context of incomplete information games, Harsanyi (1967) was the first to apply game theory to address challenges arising from incomplete information. Huang and Zhu (2020) successfully utilized it in the APT attack scenario, thereby demonstrating the applicability of game theory in resolving incomplete information issues in dynamic network confrontations. We introduce an active cybersecurity model grounded in game theory to tackle evolving network security threats, as depicted in Fig. 1.

By abstracting the interaction between the attack and the defense as a dynamic game, this model not only facilitates the effective prediction and response to attacker behavior in environments characterized by information asymmetry, but also supports the dynamic adaptation of defense strategies, thereby optimizing defense outcomes within the constraints of available resources. The four elements of SAPC establish a closed-loop defense system as a whole. Through the continuous acquisition of real-time intelligence, comprehensive threat analysis, tracking of attack vectors, and implementation of dynamic countermeasures, defenders seek to maintain an active posture in the ongoing attack–defense interaction, thereby enhancing defense efficacy and minimizing potential losses.

4.2 Intelligent threat sensing

Intelligent threat sensing constitutes the fundamental element of active cybersecurity defense,

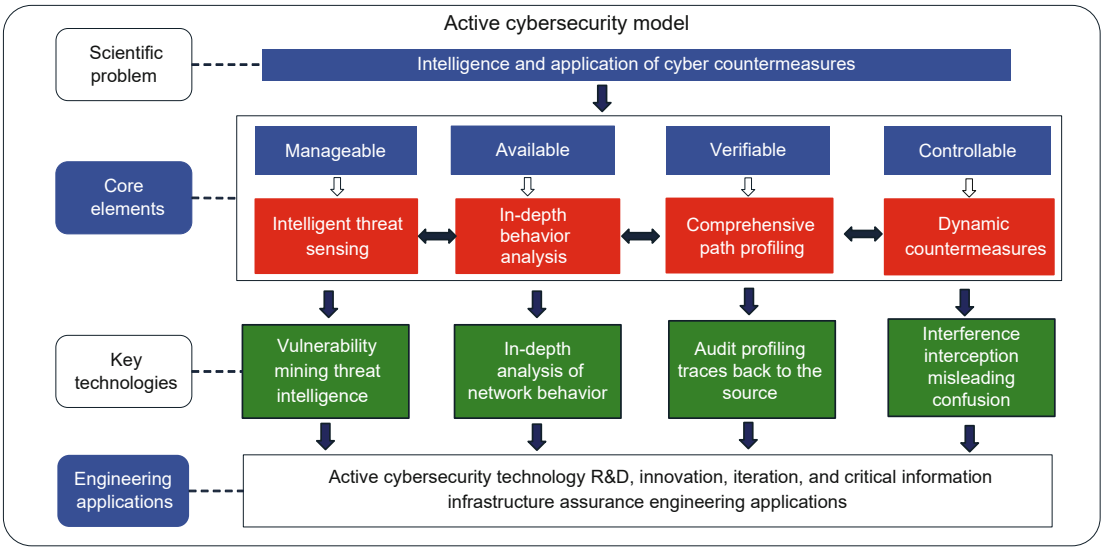

Fig. 1 Proposed sensing, analysis, profiling, and countermeasures (SAPC) model

emphasizing the early detection and prompt warning of potential threats. This enables defenders to implement timely countermeasures before attackers can infiltrate critical assets, thereby reducing the potential damage resulting from attacks. Intelligent threat sensing is grounded in two fundamental objectives. The primary objective is to facilitate an early warning, a rapid detection, and an immediate report of potential attacks. This enables the establishment of the necessary conditions for the timely implementation of countermeasures. This not only reduces the response time to threats but also significantly improves the effectiveness of security defenses. The secondary objective is to achieve an information superiority advantage. This entails the collection, analysis, and processing of security-related information more efficiently and comprehensively than those of attackers. It assists defenders in bridging the information gap with attackers, thereby ensuring that they maintain the initiative in the attack–defense dynamics. This information-driven strategy equips defenders with the capability to anticipate attacker movements, preemptively deploy defenses, and respond more flexibly to diverse attack scenarios. Basically, intelligent threat sensing concentrates on several key technologies, including penetration testing, live simulations, and vulnerability mining.

Within the context of game theory, intelligent threat sensing empowers defenders to make informed decisions by continuously gathering, processing, and analyzing intelligence, thereby alleviating the drawbacks of information asymmetry and improving response capabilities. Intelligent threat sensing, in-depth behavior analysis, comprehensive path profiling, and dynamic countermeasures work in close coordination with other defensive components, together forming an integrated defense system to achieve active cybersecurity. By providing real-time access to early warning intelligence, intelligent threat sensing guarantees that defenders can rapidly detect threats and provide crucial support to other defense components.

4.3 In-depth behavior analysis

As a crucial tool grounded in game theory, in-depth behavior analysis aids defenders in identifying potential attack indicators within intricate network attack–defense scenarios and in extracting key threat intelligence from multi-dimensional data. In the network attack–defense game, information asymmetry presents a substantial challenge for defenders. Attackers conceal their actions using techniques such as encryption, obfuscation, and camouflage. The key technologies involved in in-depth behavior analysis are the detection and analysis of multi-dimensional data such as network traffic, system logs, and other external resources.

Meanwhile, advancements in machine learning and intelligent technologies enable in-depth behavior analysis to continuously enhance its ability to assess

network activities. By leveraging a game-theoretic scenario, it adapts dynamically to changes in the network environment, allowing defenders to retain the initiative in the attack–defense game. This adaptive strategy leads to more accurate threat detection and more efficient responses to network intrusions and evolving attacks.

4.4 Comprehensive path profiling

Comprehensive path profiling uses multi-dimensional data analysis to identify the attack path and the attacker's identity. By tracing the attacker's behavioral trajectory, key nodes, and potential threats, this method supports defenders in refining both their reactive and predictive strategies. Its ultimate goal is to empower defenders to maintain control of the attack–defense game by improving response times and anticipating future threats. The key technologies in comprehensive path profiling are traceability graph analysis, log traceability, and link traceability for the identification of the attacker's action chain and key attack nodes.

By integrating game-theoretic analysis, comprehensive path profiling enables defenders to assess the attacker's strategy, potential gains, and associated risks, thus guiding the development of appropriate defense measures. Path tracing helps identify key attack nodes, allowing defenders to dynamically adjust their defenses, increase the attacker's cost, and lower the probability of successful attacks.

4.5 Dynamic countermeasures

Dynamic countermeasures are essential to active network defense. They enable real-time strategy adjustments, raise the attacker's operational cost, and may eventually force the attacker to give up. By leveraging intelligent threat sensing and comprehensive path profiling, defenders can use game-theoretic analysis to maintain a strategic edge in the ongoing attack–defense game. The key technologies in dynamic countermeasures are classified into internal and external countermeasures. The former are emergency responses, system reinforcements, real-time firewall adjustments, and device isolation, whereas the latter focus on disrupting the attacker's external support resources, including offshore springboards, command and control (C&C) servers, and anonymizing networks such as the onion router (Tor) (Zhuo et al., 2018).

Guided by game theory, dynamic countermeasures empower defenders to take control of network security by dynamically altering defense strategies, optimizing resource allocation, and disrupting the attack chain to escalate the attacker's risks and operational costs. The ultimate aim is to compel the attacker to abandon the attack, securing proactive control over the network's security position.

4.6 Game theory in the SAPC model

Active network security game theory applies game-theoretic models to assess and optimize defense strategies. It explores the strategic dynamics between defenders and attackers, assisting defenders in developing effective security plans despite information asymmetry, resource constraints, and the evolving nature of attacks.

4.6.1 Participants in the game model

The core model of active cybersecurity game theory revolves around two primary participants: the defender and the attacker. The defender, typically an organization or a network security team, aims to minimize vulnerabilities, mitigate attacks, and safeguard network assets. The defender has access to a set of defensive strategies, denoted as $S_{\mathrm{D}} = \{S_{\mathrm{s}}, S_{\mathrm{a}}, S_{\mathrm{p}}, S_{\mathrm{c}}\}$, which may include deploying firewalls, detecting intrusion, and patching vulnerabilities. These strategies require significant resource investment in terms of time, manpower, and technology. Specifically, the cost of each defensive strategy can be represented as the sum of the resource expenditures across various defensive nodes, such as monitoring systems, analysis capabilities, and response measures. The defender's strategy cost is given by the following expression:

$$C_{\mathrm{D}}(S_{\mathrm{D}}) = \sum_{i=1}^{n_{\mathrm{D}}} c_{\mathrm{D},i} R_{\mathrm{D}}^{i}, \tag{1}$$

where $c_{\mathrm{D},i}$ is the cost coefficient for resources used at the i^{th} defense node, R_{D}^{i} is the resource investment in that node, and n_{D} is the number of defense nodes.

On the other hand, the attacker is an adversarial entity focused on compromising or breaching the defender's network. The attacker's strategy set (S_{A}) encompasses various offensive tactics, such as distributed denial-of-service (DDoS) attacks, malware,

social engineering, among others, selected according to the attacker's objectives and available resources. The attacker's resource investment also incurs a cost, represented as follows:

$$C_{\mathrm{A}}(S_{\mathrm{A}}) = \sum_{i=1}^{n_{\mathrm{A}}} c_{\mathrm{A},i} R_{\mathrm{A}}^{i}, \tag{2}$$

where $c_{\mathrm{A},i}$ is the cost coefficient for resources used at the i^{th} attack path, R_{A}^{i} is the resource investment in that path, and n_{A} is the number of attack paths. The attacker aims to maximize the chance of a successful breach, often by exploiting weaknesses in the defender's defenses.

In the SAPC model, the defender is frequently confronted with the issue of incomplete information. The defender cannot immediately identify all possible attack behaviors or understand the attacker's real-time intentions. To bridge this gap, the defender must adopt strategies such as intelligent threat sensing, in-depth behavior analysis, comprehensive path profiling, and dynamic countermeasures. We begin by outlining the strategy space for these four critical elements:

1. Intelligent threat sensing involves the defender actively identifying potential threats through continuous intelligence gathering and network surveillance. This includes real-time analysis of vulnerability mining, user behavior, and external threat intelligence, aiming to detect threats before they materialize. In dynamic strategies, the strategy space for the intelligent threat sensing phase is $S_{\mathrm{s}} = \{s_1, s_2, \ldots, s_m\}$, where each s_i $(i = 1, 2, \ldots, m)$ represents a strategy, such as improving monitoring accuracy, shortening memory utilization time, or expanding intelligence collection, and m denotes the total number of strategies available in intelligent threat sensing phase.

2. In-depth behavior analysis uses multi-dimensional data analysis to thoroughly assess the attacker's behavior and detect anomalous activities or potential attack paths. The strategy space for this phase is $S_{\mathrm{a}} = \{a_1, a_2, \ldots, a_\eta\}$, where each a_i $(i = 1, 2, \ldots, \eta)$ denotes a specific behavior, such as comprehensive data analysis, machine learning applications, and association analysis, and η indicates the total number of strategies available in in-depth behavior phase.

3. Comprehensive path profiling involves real-time detection to track the attack path, identifying the source and trajectory of the attacker. The strategy space for this phase is $S_{\mathrm{p}} = \{p_1, p_2, \ldots, p_v\}$, where each p_i $(i = 1, 2, \ldots, v)$ represents a tracking method, such as deploying IDSs, analyzing logs, or utilizing path-tracking techniques, and v denotes the total number of strategies available for comprehensive path profiling.

4. Dynamic countermeasures require the defender to take immediate action upon confirming the attack. The strategy space for this phase is $S_{\mathrm{c}} = \{c_1, c_2, \ldots, c_u\}$, where each c_i $(i = 1, 2, \ldots, u)$ denotes a defensive measure, such as traffic cleaning, vulnerability patching, or blocking malicious Internet protocol (IP) addresses, and u refers to the number of strategies available for dynamic countermeasures. Additionally, this phase includes reverse induction to gain further insight into the attacker's strategy, enhancing the defender's ability to respond dynamically and effectively.

The attacker's strategy space is defined by the set of possible attack behaviors, constrained by choices such as attack paths, camouflage techniques, attack intensities, and other tactics. The attacker's strategy space is denoted as $S_{\mathrm{A}} = \{A_1, A_2, \ldots, A_w\}$, where each A_i $(i = 1, 2, \ldots, w)$ refers to a specific attack behavior, such as covert actions, path alterations, or the use of proxies, and w denotes the total number of strategies possessed by the attacker.

4.6.2 Synergy between global and local games

In active cybersecurity game theory, the game process is split into two levels: the global game which involves cross-domain collaboration and formulating overarching defense strategies, and the local game which concentrates on optimizing defense at individual nodes or systems. This structure mirrors the global and the local games in economics, where the former highlights the interdependence of strategies and the latter focuses on maximizing local interests. The four components of SAPC work together to enhance defense capabilities in the global game while addressing real-time response and local system defense in the local game.

The iterative relationship between the local and global games is outlined in Algorithm 1, where T denotes the maximum number of games. In the local game, both the defender and the attacker select strategies and update their local information based on their payoffs. The global game follows,

calculating global payoffs and prompting the parties to update their belief distributions. When the equilibrium is achieved, optimal strategies for both the defender and the attacker are determined. This model demonstrates the mutual influence of the local and global games, whereby the local game informs the global game, and feedback from the global game refines future iterations of the local game, ensuring continuous strategic optimization.

Algorithm 1 Active cybersecurity game algorithm

Input: R_{D}, S_{D}, R_{A}, S_{A}, T
Output: S_{D}^{*}, S_{A}^{*}
initialize $t=1$
while $t \leq T$ **do**
  for each $\theta_{\mathrm{D}} \in S_{\mathrm{D}}$ and $\theta_{\mathrm{A}} \in S_{\mathrm{A}}$ **do**
    $U_{\mathrm{D}}^{\mathrm{local}}\left(\theta_{\mathrm{D}}, S_{\mathrm{D}}, S_{\mathrm{A}}, R_{\mathrm{D}}\right)$
    $U_{\mathrm{A}}^{\mathrm{local}}\left(\theta_{\mathrm{A}}, S_{\mathrm{A}}, S_{\mathrm{D}}, R_{\mathrm{A}}\right)$
  end for
  $U_{\mathrm{D}}^{\mathrm{global}}(\theta_{\mathrm{D}}, S_{\mathrm{D}}, S_{\mathrm{A}}, R_{\mathrm{D}})$
  $U_{\mathrm{A}}^{\mathrm{global}}(\theta_{\mathrm{A}}, S_{\mathrm{A}}, S_{\mathrm{D}}, R_{\mathrm{A}})$
  $P\left(\theta_{\mathrm{A}} \mid S_{\mathrm{A}}\right)=\frac{P\left(S_{\mathrm{A}} \mid \theta_{\mathrm{A}}\right) P\left(\theta_{\mathrm{A}}\right)}{\sum_{\theta_{\mathrm{A}}^{\prime} \in \Theta_{\mathrm{A}}} P\left(S_{\mathrm{A}} \mid \theta_{\mathrm{A}}^{\prime}\right) P\left(\theta_{\mathrm{A}}^{\prime}\right)}$
  $P\left(\theta_{\mathrm{D}} \mid S_{\mathrm{D}}\right)=\frac{P\left(S_{\mathrm{D}} \mid \theta_{\mathrm{D}}\right) P\left(\theta_{\mathrm{D}}\right)}{\sum_{\theta_{\mathrm{D}}^{\prime} \in \Theta_{\mathrm{D}}} P\left(S_{\mathrm{D}} \mid \theta_{\mathrm{D}}^{\prime}\right) P\left(\theta_{\mathrm{D}}^{\prime}\right)}$
  $\left(S_{\mathrm{D}}^{*}, S_{\mathrm{A}}^{*}\right)=\mathrm{OptimalStrategy}\left(S_{\mathrm{D}}, S_{\mathrm{A}}\right)$
  if checkNashEquilibrium$\left(S_{\mathrm{D}}^{*}, S_{\mathrm{A}}^{*}\right)$ **then**
    break
  end if
  $t=t+1$
end while
return $S_{\mathrm{D}}^{*}, S_{\mathrm{A}}^{*}$

1. Local game: strategy optimization within distinctive stage

The local game refers to the interaction between the defender and the attacker at each stage of the security process, driven by the four components of SAPC. The process is divided into the following four key phases:

(1) Intelligent threat sensing: The defender assesses whether to allocate additional resources to enhance the capability to detect emerging threats. Meanwhile, the attacker seeks to expand the attack surface to create more opportunities for launching attacks.

(2) In-depth behavior analysis: The defender evaluates whether to invest more resources in strengthening its analytical capabilities for identifying hidden or covert attacks. In response, the attacker focuses on evading detection and increasing the complexity of audits.

(3) Comprehensive path profiling: In this phase, the defender decides whether to allocate resources to track the attacker's movements and attack paths. On the other hand, the attacker works on developing a new infrastructure to circumvent the defender's tracking efforts.

(4) Dynamic countermeasures: The defender formulates and executes a response based on previously gathered data, implementing counteractions as necessary. The attacker adapts its strategies to mitigate the effects of the defender's countermeasures, continually shifting the dynamics of the attack–defense game.

Thus, the defender's payoff function is defined as follows:

$$U_{\mathrm{D}}^{\mathrm{local}}\left(\theta_{\mathrm{D}}, S_{\mathrm{D}}, S_{\mathrm{A}}, R_{\mathrm{D}}\right)=\alpha \rho\left(\theta_{\mathrm{D}}, S_{\mathrm{D}}, S_{\mathrm{A}}\right)-\gamma R_{\mathrm{D}}, \tag{3}$$

where α represents the coefficient of successful payoff for the defender at a given stage, $\rho\left(\theta_{\mathrm{D}}, S_{\mathrm{D}}, S_{\mathrm{A}}\right)$ is the probability function of successfully achieving the objectives of intelligent threat sensing, in-depth behavior analysis, comprehensive path profiling, and dynamic countermeasures, which depends on the defender's type θ_{D}, strategy S_{D}, and the attacker's strategy S_{A}, γ denotes the coefficient of defense resource costs, and R_{D} represents the resources invested by the defender.

The attacker's payoff function is defined as follows:

$$U_{\mathrm{A}}^{\mathrm{local}}\left(\theta_{\mathrm{A}}, S_{\mathrm{A}}, S_{\mathrm{D}}, R_{\mathrm{A}}\right)=\delta \rho\left(\theta_{\mathrm{A}}, S_{\mathrm{A}}, S_{\mathrm{D}}\right)-\zeta R_{\mathrm{A}}, \tag{4}$$

where δ represents the coefficient of successful payoff for the attacker at a given stage, $\rho\left(\theta_{\mathrm{A}}, S_{\mathrm{A}}, S_{\mathrm{D}}\right)$ is the probability function of a successful attack or concealment, which depends on the attacker's type θ_{A}, strategy S_{A}, and the defender's strategy S_{D}, ζ denotes the coefficient of attack resource costs, while R_{A} represents the resources invested by the attacker in this stage.

During each round of the local game, the defender adjusts its strategies in response to the attacker's behavior. Through a continuous process of refining its understanding, minimizing information asymmetry, and progressively strengthening its defenses, the defender ultimately deploys effective countermeasures.

2. Global game: strategy optimization across multiple local games

In the global game context, both the defender and the attacker face significant challenges due to information asymmetry. The defender is unable to fully comprehend the attacker's strategies or resource allocations, and similarly, the attacker lacks a complete understanding of the defender's overall defense strategy. As a result, the defender must develop effective defense strategies by leveraging global coordination, despite having only partial information from each individual node.

In the global game, the payoff function captures the outcomes for both the attacker and the defender, accounting for strategies, resource allocations, and the asymmetry of information. For the defender, the payoff function $U_{\mathrm{D}}^{\mathrm{local}}$ which represents the local utility or payoff derived from the defense efforts at a specific node, reflects multiple dimensions, such as defense success, resource efficiency, and losses from attacks. More formally, $U_{\mathrm{D}}^{\mathrm{local}}$ is defined as follows:

$$U_{\mathrm{D}}^{\mathrm{local}}(\theta_{\mathrm{D}}, S_{\mathrm{D}}, S_{\mathrm{A}}, R_{\mathrm{D}}) = \sum_{i=1}^{n_{\mathrm{D}}} (\alpha_i \rho_{\mathrm{p}}(S_{\mathrm{D}}^i, S_{\mathrm{A}}) - \beta_i R_{\mathrm{D}}^i), \tag{5}$$

where $\rho_{\mathrm{p}}(S_{\mathrm{D}}^i, S_{\mathrm{A}})$ represents the probability of successful defense at the i^{th} node, which is influenced by the defender's strategy S_{D} and the attacker's strategy S_{A}. Here, R_{D}^i denotes the resource investment at the i^{th} node, where the increase in resources enhances the defense success rate; α_i and β_i are the weight coefficients for the nodes, reflecting the relative importance of each node to the overall defense.

The attacker's payoff function $U_{\mathrm{A}}^{\mathrm{global}}(\theta_{\mathrm{A}}, S_{\mathrm{A}}, S_{\mathrm{D}}, R_{\mathrm{A}})$ represents the overall utility or payoff that the attacker gains from executing an attack in the global context. This function typically incorporates factors such as the success rate of the attack, the damage caused, and the resources expended. The payoff is influenced not only by the local attack strategy but also by the broader impact of cross-domain attacks. More formally, the attacker's payoff function $U_{\mathrm{A}}^{\mathrm{global}}$ is defined as follows:

$$U_{\mathrm{A}}^{\mathrm{global}}(\theta_{\mathrm{A}}, S_{\mathrm{A}}, S_{\mathrm{D}}, R_{\mathrm{A}}) = \sum_{i=1}^{n_{\mathrm{A}}} (\delta_i \rho_{\mathrm{a}}(S_{\mathrm{A}}, S_{\mathrm{D}}^i) - \zeta_i R_{\mathrm{A}}^i), \tag{6}$$

where $\rho_{\mathrm{a}}(S_{\mathrm{A}}, S_{\mathrm{D}}^i)$ represents the probability that the attacker breaches the defense of the i^{th} node, which depends on both the attacker's strategy S_{A} and the defender's strategy S_{D}, R_{A}^i denotes the attacker's resource investment at the i^{th} node, with increased investment facilitating the breach of defense, and δ_i and ζ_i are the weight coefficients that reflect the priorities of different attack paths.

The global game's solution requires identifying the optimal strategy pairs for both the attacker and the defender using game-theoretic methods, such as the Nash equilibrium (Zhang LD and Hemberg, 2019). With information asymmetry in play, both players must adjust their belief distributions based on the observations of each other's strategies, which is mathematically formulated using the Bayes theorem (Yang TF et al., 2024). We denote the attacker's type as A, where the defender observes the attacker's strategy S_{A} and holds a prior belief $P(\mathrm{A})$. Applying the Bayes theorem, the defender's posterior belief is defined as follows:

$$P(\theta_{\mathrm{A}} \mid S_{\mathrm{A}}) = \frac{P(S_{\mathrm{A}} \mid \theta_{\mathrm{A}}) P(\theta_{\mathrm{A}})}{\sum_{\theta'_{\mathrm{A}} \in \Theta_{\mathrm{A}}} P(S_{\mathrm{A}} \mid \theta'_{\mathrm{A}}) P(\theta'_{\mathrm{A}})}, \tag{7}$$

where $P(\theta_{\mathrm{A}}|S_{\mathrm{A}})$ represents the defender's posterior belief, indicating the probability that the attacker's type is A, $P(S_{\mathrm{A}}|\theta_{\mathrm{A}})$ denotes the probability that the attacker adopts strategy S_{A} given the attacker's type A, and $P(\theta_{\mathrm{A}})$ is the prior probability of the attacker's type.

Similarly, when the attacker observes the defense strategy S_{D} adopted by the defender, the attacker's posterior belief is defined as follows:

$$P(\theta_{\mathrm{D}} \mid S_{\mathrm{D}}) = \frac{P(S_{\mathrm{D}} \mid \theta_{\mathrm{D}}) P(\theta_{\mathrm{D}})}{\sum_{\theta'_{\mathrm{D}} \in \Theta_{\mathrm{D}}} P(S_{\mathrm{D}} \mid \theta'_{\mathrm{D}}) P(\theta'_{\mathrm{D}})}, \tag{8}$$

where $P(\theta_{\mathrm{D}}|S_{\mathrm{D}})$ represents the posterior probability, which is the defender's updated belief about their type θ_{D} after observing the strategy S_{D}, and $P(S_{\mathrm{D}}|\theta_{\mathrm{D}})$ is the likelihood function, indicating the probability of observing strategy S_{D} given that the defender's type is θ_{D}. Further, $P(\theta_{\mathrm{D}})$ denotes the prior probability, representing the defender's initial belief about their type θ_{D} before observing any strategy, and Θ_{D} is the set of all possible defender types. The term $P(S_{\mathrm{D}}|\theta'_{\mathrm{D}})$ refers to the likelihood function for each possible defender type $\theta'_{\mathrm{D}} \in \Theta_{\mathrm{D}}$, showing the probability of observing S_{D} given a particular type. $P(\theta'_{\mathrm{D}})$ is the prior probability for each type θ'_{D}, representing the defender's belief where defender type is θ'_{D}. The denominator, $\sum_{\theta'_{\mathrm{D}} \in \Theta_{\mathrm{D}}} P(S_{\mathrm{D}}|\theta'_{\mathrm{D}}) P(\theta'_{\mathrm{D}})$, is the normalized constant, ensuring that the posterior

probabilities sum to "1" by calculating the total probability of observing S_{D} across all possible defender types, weighted by the prior belief for each type.

In a Bayesian game, the payoff functions for both participants must account for their belief distributions, which are captured as expected payoffs. The defender's expected payoff is defined as follows:

$$E\left[U_{\mathrm{D}}\right]=\sum P\left(\theta_{\mathrm{A}} \mid S_{\mathrm{A}}\right) U_{\mathrm{D}}\left(\theta_{\mathrm{D}}, S_{\mathrm{D}}, S_{\mathrm{A}}, R_{\mathrm{D}}\right), \quad (9)$$

where $\theta_{\mathrm{A}} \in \Theta_{\mathrm{A}}$ is the attacker's type, $P_{\mathrm{D}}(\theta_{\mathrm{A}}|S_{\mathrm{A}})$ is the defender's posterior belief over the attacker's type upon observing the attacker's strategy S_{A}, and $U_{\mathrm{D}}(\theta_{\mathrm{D}}, S_{\mathrm{D}}, S_{\mathrm{A}}, R_{\mathrm{D}})$ is the defender's payoff given both players' types and strategies.

The attacker's expected payoff is defined as follows:

$$E\left[U_{\mathrm{A}}\right]=\sum P\left(\theta_{\mathrm{D}} \mid S_{\mathrm{D}}\right) U_{\mathrm{A}}\left(\theta_{\mathrm{A}}, S_{\mathrm{A}}, S_{\mathrm{D}}, R_{\mathrm{A}}\right), \quad (10)$$

where $\theta_{\mathrm{D}} \in \Theta_{\mathrm{D}}$ is the defender's type, $P_{\mathrm{A}}(\theta_{\mathrm{D}}|S_{\mathrm{D}})$ is the attacker's posterior belief over the defender's type upon observing the defender's strategy S_{D}, and $U_{\mathrm{A}}(\theta_{\mathrm{A}}, S_{\mathrm{A}}, S_{\mathrm{D}}, R_{\mathrm{A}})$ denotes the attacker's payoff given both players' types and strategies.

Given the costs and benefits associated with each participant's strategies, it is important to evaluate the return on investment (ROI) for both the defender and the attacker. The ROI is a measure of the net benefit relative to the cost incurred by each participant. The ROI can be calculated as follows:

The defender's ROI is defined as

$$\mathrm{ROI}_{\mathrm{D}}=\frac{E\left[U_{\mathrm{D}}\right]-C_{\mathrm{D}}\left(S_{\mathrm{D}}\right)}{C_{\mathrm{D}}\left(S_{\mathrm{D}}\right)}, \quad (11)$$

where $E\left[U_{\mathrm{D}}\right]$ is the defender's benefit and $C_{\mathrm{D}}\left(S_{\mathrm{D}}\right)$ is the cost of the defender's strategy.

The attacker's ROI is defined as

$$\mathrm{ROI}_{\mathrm{A}}=\frac{E\left[U_{\mathrm{A}}\right]-C_{\mathrm{A}}\left(S_{\mathrm{A}}\right)}{C_{\mathrm{A}}\left(S_{\mathrm{A}}\right)}, \quad (12)$$

where $E\left[U_{\mathrm{A}}\right]$ is the attacker's benefit and $C_{\mathrm{A}}\left(S_{\mathrm{A}}\right)$ is the cost of the attacker's strategy.

Both the defender and the attacker aim to maximize their ROIs by optimizing their respective strategies. In each round of the game, the defender and the attacker adjust their strategies dynamically, based on their resources and the effectiveness of their actions.

In a global game with information asymmetry, both the defender and the attacker achieve a Nash equilibrium by selecting their respective optimal strategies, so that neither party has an incentive to change their strategy given the strategy of the other. Specifically, a Nash equilibrium occurs when the defender and the attacker choose the optimal strategies, respectively, under the condition in which both maximize their expected payoffs. Furthermore, to maximize the return on investment, both parties must choose strategies that effectively balance the defense and the attack while considering the costs of their actions, thereby finding a balance between resource efficiency and effectiveness. The Nash equilibrium is listed as follows:

$$\begin{aligned}&\left(S_{\mathrm{D}}^{*}, S_{\mathrm{A}}^{*}\right)=\mathrm{OptimalStrategy}\left(S_{\mathrm{D}}, S_{\mathrm{A}}\right)\\&\quad \text{s.t.} \quad \arg \max_{S_{\mathrm{D}}, S_{\mathrm{A}}}\left[\mathrm{ROI}_{\mathrm{A}}, \mathrm{ROI}_{\mathrm{D}}\right],\end{aligned} \quad (13)$$

where S_{D}^{*} and S_{A}^{*} denote the optimal strategies for the defender and the attacker, respectively. The expression implies that each player is choosing the strategy that maximizes their expected utility, considering the strategy of the other player.

5 Key technologies based on the SAPC model

This section explores practical case studies to illustrate how these key technologies improve defensive capabilities in complex and dynamic network environments.

5.1 Overview

As illustrated in Fig. 2, the active network security model is built on an extensive technical framework, designed to form a closed-loop defense system via proactive and intelligent technological methodologies.

The system includes four primary technical modules: vulnerability mining, traffic detection, source tracing analysis, and countermeasures. These modules complement one another, working together to offer real-time, effective protection and responses at different stages of network security threats.

The SAPC model consists of four core elements, intelligent threat sensing, in-depth behavior analysis, comprehensive path profiling, and dynamic countermeasures, each serving a distinct function within the

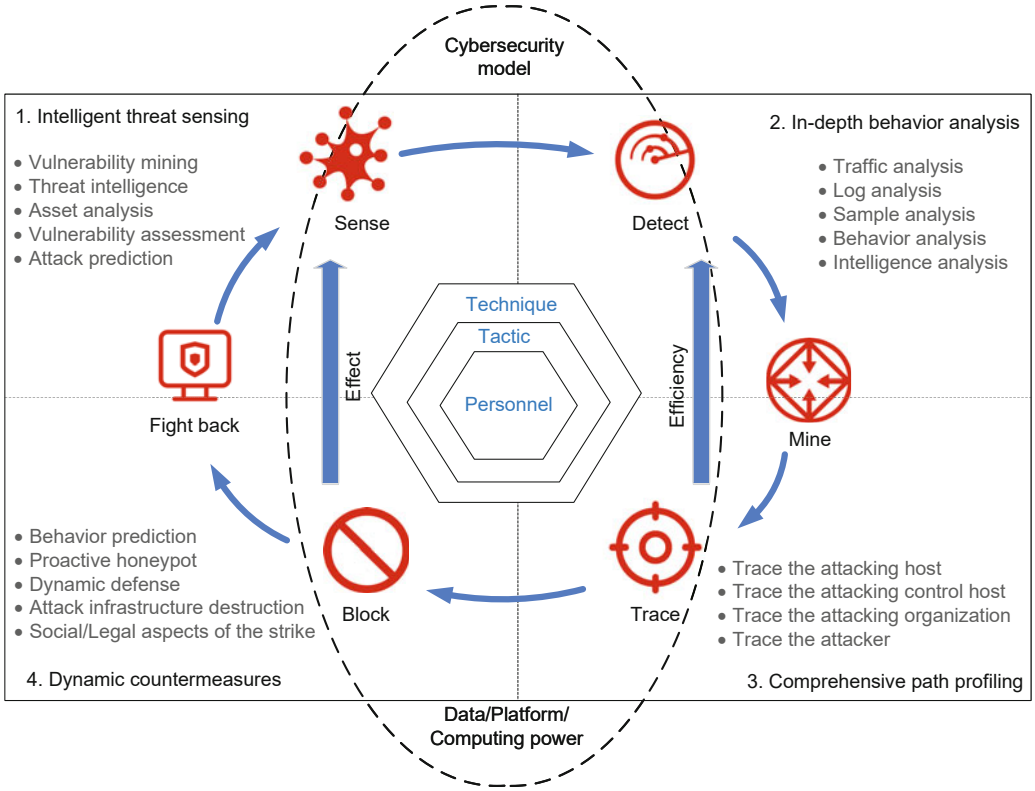

Fig. 2 Technical framework of the sensing, analysis, profiling, and countermeasures (SAPC) model

system. These elements are interconnected, forming a cyclical structure that facilitates continuous feedback. "Intelligent threat sensing" initiates the cycle by proactively identifying vulnerabilities and predicting potential threats, informing "in-depth behavior analysis" to detect anomalies and malicious behaviors. Upon detection, "in-depth behavior analysis" forwards suspicious activity data to "comprehensive path profiling," which reconstructs attack paths and traces origins, thereby providing actionable intelligence for "dynamic countermeasures" to deploy targeted countermeasures. Crucially, postresponse feedback of "dynamic countermeasures" refines the accuracy of threat predictions, anomaly detection models, and path-tracing algorithms, creating a self-reinforcing loop. This cyclical interaction is further enhanced by robust data integration, high-performance computing, and a unified platform, ensuring seamless information flow and real-time optimization. This interconnectedness results in a highly coordinated system, where the logical relationships and interactions between the elements are emphasized. Moreover, the cyclical structure supports a closed-loop mechanism, enabling the network security system to respond swiftly, optimize in real time, and adapt continuously to emerging threats.

Intelligent threat sensing, as the cornerstone of active network security, plays a vital role in identifying potential vulnerabilities and risks. Prior to an attack, intelligent sensing continuously monitors and analyzes network assets to detect weaknesses that could be exploited by adversaries. The goal of intelligent sensing is to use advanced threat detection and prediction technologies to accurately assess the intentions, behavioral patterns, and potential attack paths of prospective attackers. This proactive approach facilitates the early deployment of defense strategies, with a primary focus on identifying system vulnerabilities before an attack is executed to prevent its success.

In-depth behavior analysis, a crucial component of traffic detection, utilizes traffic monitoring and data analysis techniques to identify abnormal activities or malicious behaviors in real time. The primary goal of this analysis is to predict and identify the actions of attackers by examining multi-dimensional

data, revealing attack paths, and providing critical intelligence to inform defense strategies. Upon detecting traffic anomalies, the system rapidly analyzes data packets to determine the presence of potential attack behaviors, such as DDoS attacks or malware propagation. This detailed analysis not only aids in detecting attack patterns but also generates valuable data that support intelligent perception, refining vulnerability mining strategies. Moreover, new vulnerabilities identified by intelligent perception can guide adjustments in diagnostic technologies, improving the accuracy of traffic detection and expanding the system's ability to identify a wider array of attacks.

Comprehensive path profiling, a key component of source tracing analysis, tracks the source and path of an attack after it occurs, providing a detailed understanding of the attack's process and methodology. This approach not only supports defense mechanisms during ongoing attacks but also provides valuable insights and data that can inform future protection strategies. The primary objective of this process is to trace the attack path, reconstruct the attack chain, and thoroughly analyze the attacker's behavioral trajectory, enabling defenders to implement more targeted countermeasures, particularly against APT attacks. By tracing the attack, the system can retrospectively verify abnormal behaviors identified during earlier analysis, pinpoint the attack chain, and refine the countermeasure strategy. In addition, the dynamic feedback from this process enhances other elements, ensuring that tasks such as vulnerability mining and traffic detection are better aligned with emerging threats.

Dynamic countermeasures, a pivotal element of network defense, play a crucial role in mitigating the escalation of attacks by providing rapid responses and remediation. The primary objective of countermeasures is to force attackers to abandon their efforts or reduce the effectiveness of their attacks by raising their associated costs. These countermeasures not only serve as immediate responses to active threats but also enable the adaptation of defense strategies based on the insights derived from source tracing analysis, thereby enhancing both the accuracy and timeliness of defensive actions. The effectiveness of countermeasures is fed back into each component of system-intelligent perception, in-depth analysis, and dynamic path profiling, continuously refining and strengthening the overall defense capabilities. Through a closed-loop feedback mechanism, countermeasures not only address current threats but also promote the system's ability to learn from and adapt to emerging attack patterns, ensuring sustained and effective protection against future attacks.

To achieve a seamless closed-loop collaborative operation, the active network security framework depends on a combination of robust data, computing power, and platforms to enhance the system's overall performance. Data form the foundation of active network security, as the system collects and analyzes vast amounts of network traffic, log data, external threat intelligence, and other relevant information. This enables continuous, real-time monitoring of the network environment, facilitating the identification of potential threats and attack patterns. These data further support critical tasks such as vulnerability mining, traffic detection, and source tracing analysis, improving the accuracy of threat detection and response. Computing power is indispensable to active network security, especially when handling complex tasks such as large-scale data analysis, behavioral analysis, and machine learning. Powerful computational resources are essential to ensure rapid responses to emerging threats. Through the use of large-scale parallel computing and efficient data processing, the active security framework achieves high efficiency in real-time detection and automatic responses. The platform, as the underlying infrastructure, supports the integration of data, computing power, and various security technologies to offer multilayered protection. An effective security platform provides necessary services such as data storage, computational resources, and algorithmic support, while managing and optimizing the operation of various security components tailored to different threat scenarios. The platform ensures intelligent coordination and scheduling, boosting the adaptability and flexibility of the entire system.

Through the utilization of these resources, a network security large language model can be developed and deployed, demonstrating advanced intelligence and automation when confronting complex network threats. This model integrates sophisticated technologies such as big data, AI, and deep learning to bolster defense mechanisms. The network security model not only strengthens the overall defensive infrastructure but also continually refines its capacity to detect new and emerging threats through

expansive data-driven learning, fostering a dynamic "self-evolution" mechanism. This approach ensures that the active network security framework remains adaptable and effective against the ever-changing nature of attack strategies.

The four core components within the active network security framework are interdependent, functioning through a cyclical feedback mechanism that ensures continuous system optimization and real-time responsiveness. In this integrated system, the collaboration of data, computing power, and platforms lays a solid foundation for the network security model's operation and evolution. The resulting intelligent, automated defense system enhances both the speed and the accuracy of network security operations while facilitating ongoing learning and adaptation. Through continuous refinement, the system can effectively address emerging threats, maintaining robust defense capabilities at all times.

5.2 Vulnerability mining

Intelligent threat sensing, a key component of the active network security model, aims to precisely identify the intentions, behavior patterns, and potential attack paths of attackers through advanced threat detection and predictive technologies. This allows for the proactive deployment of defense strategies prior to an attack (Wang J et al., 2023).

To achieve this, it is essential to identify vulnerabilities in the target system at an early stage and block critical entry points that could be exploited by attackers. Fig. 3 depicts the interaction between attackers and defenders during the vulnerability mining process. Both focus on the same target, with attackers seeking to uncover as many vulnerabilities as possible to construct effective attack paths. Meanwhile, defenders conduct in-depth analyses of the system's security state, allowing them to actively block potential threat entry points and constrain attacker actions at an early stage. A key element of this process is the implementation of intelligent threat sensing, which enables the prediction of attackers' potential paths and targets in advance.

5.2.1 Background

Traditional vulnerability mining technologies adopt a narrow perspective. Despite advancements in reinforcement learning (Kröse, 1995), AI (Pouyanfar et al., 2019), and large language models (Zhou et al., 2018), achieving intelligent sensing remains a challenge. This challenge arises because vulnerability mining primarily relies on the defender's experience and summaries of known attacker exploits. This approach remains rooted in passive defense, lacking active consideration of unknown threats and exploitable vulnerabilities from the attacker's perspective. As a result, defenders face difficulties in anticipating new strategies and attack paths, thereby leaving security gaps. To achieve comprehensive perception of unknown threats, it is essential to break perspective limitations and redefine vulnerability mining through the attacker's lens, incorporating dynamic game theory.

5.2.2 Proposed technology

To overcome the limitations of traditional experience-based vulnerability mining, we propose an active network security vulnerability mining approach that integrates game theory. This approach focuses on the attacker's perspective, leveraging dynamic games to uncover unknown vulnerabilities and addressing the limitations of traditional technologies in predicting emerging threats and potential exploits. The technical framework includes target identification, model training, vulnerability generation, and vulnerability combination, covering the entire lifecycle of vulnerability mining to enable active perception and dynamic response.

Target combining: In this stage, the technology constructs a multiscenario confrontation model and a vulnerability mining framework by comprehensively analyzing historical vulnerability data, APT attack indicators (Hossain et al., 2017), and changes in target assets. By incorporating game theory, vulnerability mining shifts to the attacker's perspective, simulating their potential strategies and attack paths. For example, attackers may target configuration vulnerabilities in newly added network assets or infiltrate key nodes along the APT attack chain to gain permissions through phishing (Zhang Y et al., 2007), unpatched vulnerabilities (Wang XD et al., 2019), or social engineering (Heartfield and Loukas, 2016). Attackers may also exploit supply chain vulnerabilities or third-party dependencies to conduct indirect attacks (Liang et al., 2023). By simulating these strategies, the technology actively identifies potential weaknesses, overcoming the limitations of traditional empirical paradigms and providing

defenders with more precise protection targets.

Model training: In this stage, the technology applies large language models and reinforcement learning to develop a dynamic vulnerability prediction framework. This framework enables defenders to track the evolution of attack strategies through iterative optimization and adaptive learning, improving the prediction of unknown vulnerabilities. By moving beyond reliance on historical data, it facilitates a more comprehensive identification of potential threats.

Vulnerability generation: In this stage, this technology integrates fuzzing (Sutton et al., 2007) to simulate potential exploitation methods used by attackers, assess their system interactions, and evaluate their impact. A standardized vulnerability management database quantifies the cause, risk, exploitation conditions, and repair costs for each vulnerability, providing preanalysis data for attack path evaluation during the vulnerability combination stage.

Vulnerability combination: In this stage, the synergistic effects of individual vulnerabilities are analyzed to construct a multivulnerability linkage model and a comprehensive attack chain (Xu K et al., 2024). Game theory is applied to simulate attackers' optimal path selection in complex systems. For instance, attackers might escalate their privileges incrementally via multiple vulnerabilities, ultimately gaining control of critical assets. Evaluating these paths enables defenders to identify high-risk vulnerability chains, prioritize core repairs, devise efficient protection strategies, and enhance overall system security.

When integrated with the active network security model and game theory, the introduction of dynamic games becomes highly significant. The core concept lies in the Nash equilibrium of the game. When both the attacker and the defender select their optimal strategies, the game tends toward the Nash equilibrium, making the attacker's behavior predictable. In the context of vulnerability exploitation, to maximize their benefits, the attacker must choose the optimal attack path, such as prioritizing key high-value modules or resources. If the attacker deviates from the optimal strategy, their benefits will be significantly reduced, placing them at a disadvantage in the game. Therefore, vulnerability mining technology, combined with game theory, can identify high-risk vulnerabilities in advance and enable proactive defensive measures by deducing the attacker's optimal path.

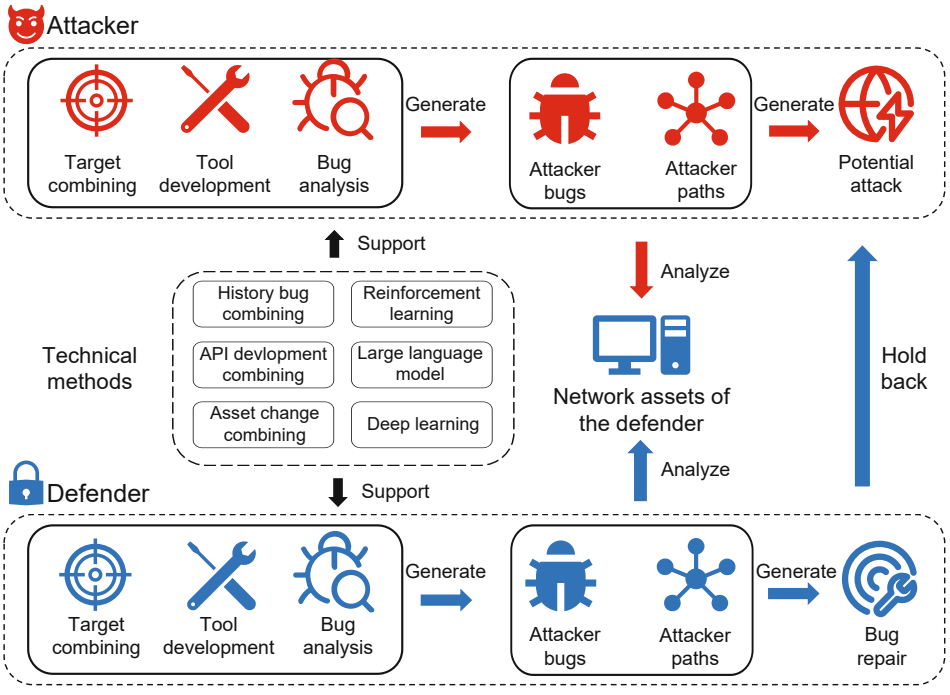

Fig. 3 Vulnerability mining technology based on active cybersecurity with game theory

5.2.3 Case study

The following example demonstrates the application of game theory in vulnerability mining technology within active network security. The attacker (A) and the defender (D) engage in a competitive scenario through vulnerability mining and repair activities. The attacker aims to exploit undiscovered or unrepaired vulnerabilities and construct effective attack paths to maximize their benefits. In contrast, the defender seeks to address vulnerabilities identified by the attacker through mining and repair while minimizing potential system vulnerabilities that remain undiscovered by the attacker.

In the vulnerability mining game, the goal of the defender is to identify the attack path, infer the attack intention, and then cut off the attack chain. Its strategy space is $S_{\mathrm{s}} = \{S_{\mathrm{IC}}, S_{\mathrm{VE}}, S_{\mathrm{VR}}\}$, where S_{IC} represents the collection of public attacker intelligence, such as historical vulnerability data and APT equipment development clues, to prepare for the next target selection, S_{VE} includes the selection of asset and vulnerability types, the development of corresponding vulnerability mining tools, and mining vulnerability, and S_{VR} represents the repair of vulnerabilities along the attack path in order of their values. The defender mines vulnerabilities by analyzing the attacker's goals and manages and repairs vulnerabilities with limited resources. The profit function is defined as follows:

$$U_{\mathrm{D}}^{\mathrm{S}} = \alpha_1 \rho_{\mathrm{Repair}}(\theta_{\mathrm{D}}, \{S_{\mathrm{IC}}, S_{\mathrm{VE}}, S_{\mathrm{VR}}\}, S_{\mathrm{A}}^{\mathrm{S}}) - \gamma_1 R_{\mathrm{D}}^{\mathrm{S}}, \tag{14}$$

where $R_{\mathrm{D}}^{\mathrm{S}}$ represents the resources consumed by the defender, ρ_{Repair} represents the probability of successfully discovering or repairing a vulnerability, and α_1 and γ_1 are weights.

Correspondingly, the attacker aims to evade detection and prolong the attack's duration by obscuring the attack path and concealing behavioral characteristics. Its strategy space is $S_{\mathrm{A}}^{\mathrm{S}} = \{a_{\mathrm{IC}}, a_{\mathrm{VM}}\}$, where a_{IC} represents the collection of public attacker intelligence and combination of the existing technology to prepare for the next target selection, and a_{VM} refers to the selection of assets, the development of corresponding vulnerability mining tools, and the mining of vulnerabilities. The attacker's goal is to discover high-value vulnerabilities in the defender's network assets. The profit function is defined as follows:

$$U_{\mathrm{A}}^{\mathrm{S}} = \delta_1 \rho_{\mathrm{Mine}}(\theta_{\mathrm{D}}, \{a_{\mathrm{IC}}, a_{\mathrm{VM}}\}, S_{\mathrm{D}}^{\mathrm{S}}) - \zeta_1 R_{\mathrm{A}}^{\mathrm{S}}, \tag{15}$$

where $R_{\mathrm{A}}^{\mathrm{S}}$ represents the resources consumed by the attacker, ρ_{Mine} represents the probability of successfully mining a vulnerability, and δ_1 and ζ_1 are weights.

In our previous research work (Wang D et al., 2020), we proposed a novel approach, BCFuzzer, for common gateway interface (CGI) vulnerability detection on embedded devices. This method leverages a feedback-driven lazy input model and selective external function tracking to automate code path exploration and efficiently identify vulnerabilities. Experimental results on real embedded CGI programs demonstrate that the proposed approach can detect more unknown vulnerabilities more quickly, with a reduction in detection time of approximately 50% and a nearly 2.5 times greater number of discovered vulnerabilities compared to existing methods.

In 2016, one year before the official disclosure of the Microsoft Office remote code execution vulnerability (CVE-2017-11882), our proprietary active cybersecurity-based vulnerability discovery system successfully identified and analyzed the attack patterns associated with this critical exploit 14 months prior to industry recognition. Leveraging adversary behavior-profiling techniques and in-depth component interaction analysis, our system detected anomalous document structure patterns during the vulnerability's latent phase, initiating real-time threat alerts through automated exploit chain mapping. This breakthrough enabled partner organizations to implement enterprise-wide document security gateways and deploy memory protection protocols at scale, completing critical infrastructure hardening a full year prior to the vulnerability's public disclosure and subsequent weaponization in global cyberattacks.

5.2.4 Summary

This technology can effectively identify potential risks before an attack occurs and adjust defense strategies in real time, thereby preventing attackers from exploiting zero-day vulnerabilities to gain unauthorized access and steal data. Additionally, this approach provides a comprehensive view of the attack

path and offers targeted repair suggestions and optimization solutions for the defender. In practice, this active vulnerability mining technology, based on game theory, significantly enhances the defender's security resilience, reduces the likelihood of security incidents, and facilitates the transition of network security from passive defense to active response.

5.3 Traffic detection

As a key component of SAPC, the objective of "in-depth behavior analysis" is to identify and predict attack actions while uncovering the attack path through comprehensive analysis of multi-dimensional data (Kumar and Agrawal, 2023), thus providing essential intelligence and strategic support for defenders. Achieving this goal involves technical methods such as traffic audit evasion (Li PY et al., 2022), which enable thorough analysis of traffic data in highly adversarial environments. As a result, "in-depth behavior analysis" focuses on leveraging active traffic detection technologies, combining historical attribute analysis of attack traffic with real-time updated feature intelligence and thereafter dynamically adjusting defense strategies in response to evolving attacker behaviors.

5.3.1 Background

Traditional traffic detection technologies (Jiang JC et al., 2000) identify specific attacks through rule-based systems. However, rapid technological advancements over the past decade have significantly increased the scale of networks and the number of applications (Arjunan, 2024), leading to an increasingly complex network environment. Nowadays, attackers can exploit a variety of both old and new attack variants to bypass these detection rules and conduct intrusions (He K et al., 2023). To address the high maintenance costs of traditional rule-based IDSs and improve their effectiveness and adaptability to modern traffic, machine learning and deep learning-based traffic detection technologies (Xia et al., 2001) can extract valuable insights from real-time network traffic instances, facilitating the detection of previously unknown traffic. Compared to rule-based methods, these approaches offer superior detection performance and enhanced adaptability to changing environments. However, these solutions face significant challenges when handling intrusion traffic variants, such as adversarial examples (AEs) (Liu QX et al., 2021). AEs can exploit discontinuities in input–output mapping hidden in deep neural networks, targeting decision boundary flaws to evade machine learning and deep learning-based detection.

5.3.2 Proposed technology

To address the challenge posed by traditional traffic detection technologies in handling intrusion traffic variants, we propose an anti-obfuscation network intrusion detection technology specifically designed for AEs, as shown in Fig. 4. This innovative technology adopts the perspective of the attacker. Initially, it simulates the knowledge extraction and detection logic of the original deep learning detection system model through multi-dimensional feature modeling based on black-box model migration (Kheddar et al., 2023). Subsequently, high-quality traffic AEs are generated using generative AI techniques (Xiao CW et al., 2018) or by introducing perturbations to the original samples (Goodfellow et al., 2014). The migrated model of the original detection system is then applied to identify flaws and bypass traffic audits. Finally, the original detection system is dynamically optimized and defect-corrected based on the characteristics of the malicious traffic (He K et al., 2023), enhancing its detection capabilities for traffic AEs. The entire training process integrates game theory, facilitating an in-depth analysis of malicious traffic characteristics and adversarial detection.

The framework of this technology comprises target establishment, attack obfuscation, and adversarial detection.

Target establishment: The objectives of the detection system are defined to guide the design of targeted attacks. This is accomplished through the systematic extraction of the statistical characteristics and behavioral attributes of the original attack traffic, enabling a thorough understanding of its feature distribution (Zhang B et al., 2023). This process involves multi-dimensional feature modeling based on time-series characteristics, packet size distribution, protocol usage patterns, and communication behavior, defining the distinctive traffic patterns of the attack. Additionally, the original detection system is reconstructed to simulate its detection logic within the network environment, reflecting the decision-making processes and vulnerabilities inherent in this

system. This system is a white-box system as its detection logics and potential weaknesses are revealed, offering direct insights for designing subsequent targeted attacks.

Attack obfuscation: AEs are generated based on the target system to evade its detection. By exploiting the highly discontinuous decision boundary flaws in the input–output mapping in this system, traffic AEs are crafted to disrupt its classification capability, complicating the accurate identification or detection of abnormal traffic. As a result, the attacker can bypass the audit process from this system and ultimately execute a covert attack scenario. AEs not only highlight decision blind spots in the original detection system but also reveal the latter's vulnerability to specific perturbed traffic. To generate AEs, our previous work (Ding et al., 2021) proposed an efficient black-box adversarial attack generation method based on transferability. By enhancing the transferability and adversarial vector of the score-based black-box attack method, attack query efficiency is maximized, reducing the number of queries required. Simultaneously, the substitution model and optimization objective function are utilized to refine the adversarial attack generation process. Furthermore, Liu XL et al. (2020) introduced a weighted sampling technique for AE generation, balancing the number of distorted samples and their respective weights to enhance attack effectiveness. A denoising method is incorporated into the loss function to improve the stealth of the attack, ultimately achieving low-noise, high-robustness AE generation.

Adversarial detection: To detect traffic AEs, the detection capability of the original system is enhanced through in-depth analysis, aiming to identify and investigate their potential features. Concurrently, the robustness of this detection system is progressively improved by integrating dynamic optimization and the correction of decision boundary flaws in its detection logic. Throughout this process, the decision boundary is iteratively optimized to ensure that detection logic for normal traffic remains unaffected. As a result, the detection system's ability to identify traffic AEs is significantly strengthened, along with its adaptability to recognize complex traffic patterns. To detect AEs, our research work (He JP et al., 2022) proposed a set of anti-obfuscation intrusion detection training methods based on adversarial training. This approach utilizes traffic AEs, original attack traffic, and original normal traffic generated by the adversarial network to update the targeted detection model, ultimately achieving comprehensive adaptation and effective detection of traffic AE representations.

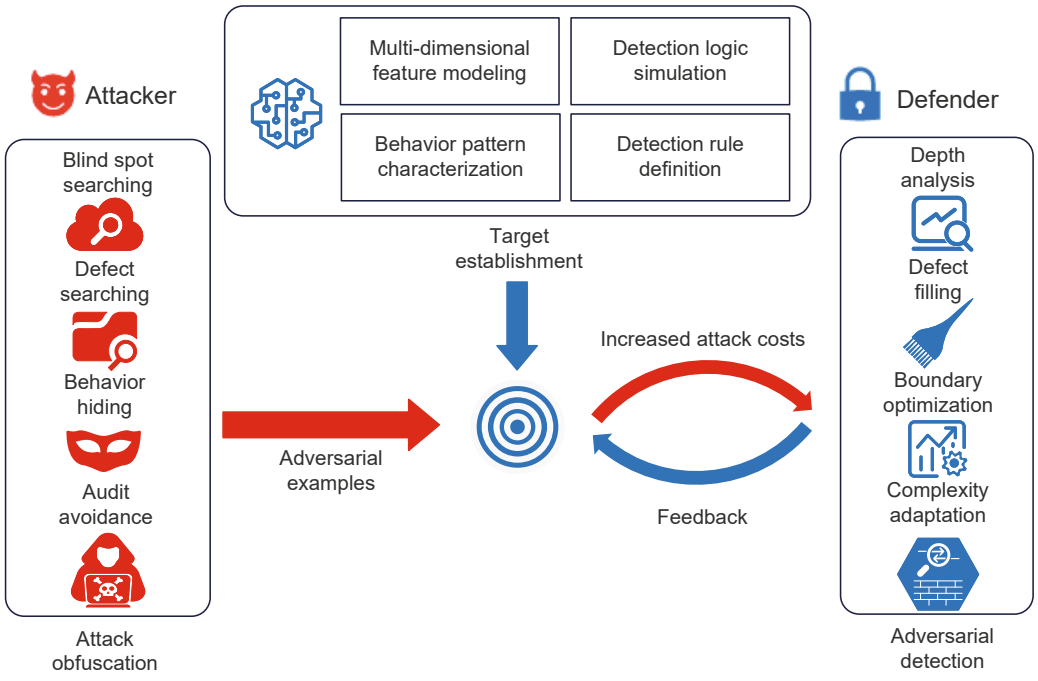

Fig. 4 Traffic detection technology based on active cybersecurity with game theory

5.3.3 Case study

Taking a previous work (He JP et al., 2022) as an example, the defender applies an IDS to detect malicious AEs generated by the attacker. The strategy space of the defender is $S_{\mathrm{a}}=\{t_{\mathrm{AT}}\}$, where t_{AT} is the "adversarial training" that utilizes the obtained AE "samples" to enhance the IDS model. The strategy space of the attacker is $S_{\mathrm{A}}=\{t_{\mathrm{DS}}\}$, where t_{DS} is "defect searching", which exploits the discontinuous decision boundary "defects" in the detection model. Thus, the profit function of the defender is defined as

$$U_{\mathrm{D}}^{\mathrm{A}}=\alpha_2\rho_{\mathrm{Identify}}(\mathrm{IDS},t_{\mathrm{AT}},t_{\mathrm{DS}})-\gamma_2 n_{\mathrm{sam}}, \quad (16)$$

where ρ_{Identify} is the probability function of successful identification by the IDS, α_2 and γ_2 are weights, and n_{sam} is the number of AE samples. The profit function of the attacker is defined as

$$U_{\mathrm{A}}^{\mathrm{A}}=\delta_2\rho_{\mathrm{Conceal}}(\mathrm{AE},t_{\mathrm{DS}},t_{\mathrm{AT}})-\zeta_2 n_{\mathrm{def}}, \quad (17)$$

where ρ_{Conceal} is the probability function of successful concealment by the traffic AEs, δ_2 and ζ_2 are weights, and n_{def} is the number of decision boundary defects.

Under the confrontation between the attacker and defender in a previous work (He JP et al., 2022), four types of adversarial attacks (DDoS, denial-of-service (DoS), bruteforce, and infiltration) based on the CICIDS2017 traffic dataset are generated with the Fréchet inception distance of 20–50. These attacks can bypass traditional machine learning models; thus, their detection rates are 0.9%–63%. Compared with other works (Yan et al., 2019; Lin et al., 2022) that concentrated on attack generation and only played the role of attackers, our work also improved IDS establishment, including concrete and systematic strategies for cybersecurity defenders. As a result, it trained an enhanced IDS that possesses on average 98% detection rates in coping with these generated adversarial attacks together with the original attacks.

5.3.4 Summary

The application of game theory in this traffic detection technology is apparent in both the attack obfuscation and adversarial detection parts. The attacker continually searches for new discontinuity mapping defects to generate traffic AEs that evade detection, while the defender enhances its resistance to them through adversarial training. The interaction between traffic audit evasion and anti-obfuscation detection creates a game-like confrontation. Through dynamic Nash equilibrium analysis, this process allows the detector to progressively address the mapping defects in the attack–defense game, increasing the difficulty for the attacker in identifying new vulnerabilities. Ultimately, this leads to comprehensive analysis and resolution of the traffic AEs by the defender.

Overall, this technology facilitates a comprehensive analysis of traffic characteristics in a highly adversarial environment by resisting the traffic audit evasion strategies of the attackers. It lies in the concept of "in-depth behavior analysis" for active network security. Besides AE detection, in-depth behavior analysis is utilized in application layer encryption feature changes (Li HH et al., 2021) and information hiding techniques (Kheddar et al., 2024) to strengthen the reliability and security of the detection system.

5.4 Attack traceback

Comprehensive path profiling tracks attack paths, reconstructs attack chains, and analyzes attacker behaviors. This enables defenders to implement precise countermeasures, particularly against APTs. Using game theory optimization, we analyze how defenders can optimize their strategies during traceback by considering both attacker and defender objectives and techniques.

5.4.1 Background

As shown in Fig. 5, attackers and defenders form a noncooperative game in network confrontation. Attackers apply IP proxies (Bocovich et al., 2024), host pivoting, anonymous networks (Oh et al., 2022; Chao et al., 2024), and code obfuscation to evade tracing. Defenders counter with attack path perception (Kim et al., 2022; Xiong et al., 2022) and attack intent reasoning (Zhao et al., 2020; Alsaheel et al., 2021; Zengy et al., 2022) techniques. Through analyzing attack patterns, discovering hidden nodes, and predicting attack behaviors, defenders can optimize their traceback strategies to improve attack attribution.

Game theory guides defenders in optimizing

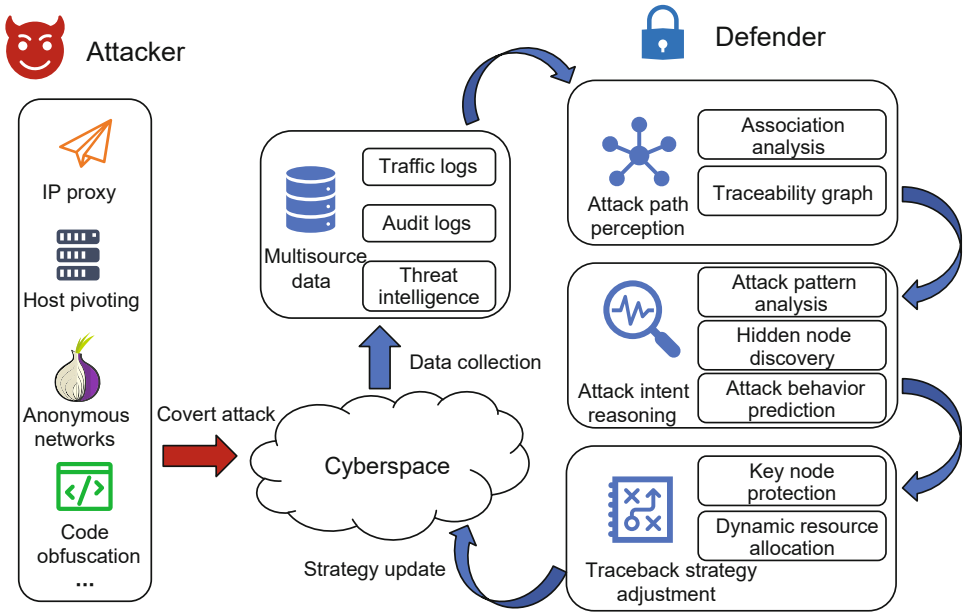

Fig. 5 Attack traceback technology based on active cybersecurity with game theory

resource allocation and traceback strategies under resource constraints, aiming to reach a Nash equilibrium (Nash, 2002; Manshaei et al., 2013) with attackers. At equilibrium, attackers choose optimal attack paths based on cost–benefit analysis, while defenders select traceback strategies based on resources and threat assessment. This enables defenders to model attack patterns, predict future moves, and actively disrupt attack chains. Through game theory, defenders can continuously adapt their strategies in response to attacker behaviors.

5.4.2 Proposed technology

In attack traceback, defenders aim to identify attack paths and intentions while optimizing their resource allocation and traceback strategies. This is achieved through three key processes: attack path perception, attack intent reasoning, and strategy adjustment.

Attack path perception: The defender builds a traceback dataset from network traffic logs, audit logs, and threat intelligence. These data undergo feature extraction through static and dynamic analysis, focusing on API call sequences (Xiong et al., 2022), process interactions, and network patterns. The defender then constructs an attack path graph (Hossain et al., 2017; Milajerdi et al., 2019), where nodes represent entities (processes, files, or IP addresses), and edges show behavioral dependencies. By incorporating historical attack patterns (Satvat et al., 2021; Zeng et al., 2021), the defender can identify key nodes and predict potential attack chains.

Attack intent reasoning: Knowledge graphs (Jia et al., 2018) enhance traceback by linking current attacks to known APT groups through historical attack data and threat intelligence. By matching behavior patterns and tools with specific APT groups' known characteristics, defenders can identify attackers and predict their intentions, enabling targeted defensive responses.

Traceback strategy adjustment: Deep learning models enable defenders to uncover hidden patterns and critical nodes in attack paths. By incorporating dynamic game optimization, defenders can continuously adjust their strategies and allocate resources based on risk priority, focusing on high-value nodes to maximize the traceback efficiency.

5.4.3 Case study

In the game process of attack traceback confrontation, the defender's objective is to identify attack paths, infer attack intentions, and disrupt the attack chain. The strategy space is defined as $S_{\mathrm{D}}^{\mathrm{T}} = \{p_{\mathrm{PP}}, p_{\mathrm{IR}}, p_{\mathrm{SA}}\}$. Here, p_{PP} represents attack path perception, which integrates traffic analysis, audit logs, and threat intelligence to construct

an attack path traceback graph and identify critical nodes in the attack chain; p_{IR} represents attack intent reasoning, which correlates historical attack patterns with current attack features based on a knowledge graph to infer attack intentions and predict the attacker's next actions; p_{SA} represents traceback strategy adjustment, which optimizes traceback strategies by prioritizing resource allocation to high-value nodes and precisely intervening in the attack chain.

To optimize the success rate of attack traceback, the defender seeks to accurately identify attack paths and infer attack intentions, while simultaneously enhancing efficiency within the constraints of limited resources. The payoff function is defined as

$$U_{\mathrm{D}}^{\mathrm{T}}=\alpha_3\rho_{\mathrm{Traceback}}\left(\theta_{\mathrm{D}},\{p_{\mathrm{PP}},p_{\mathrm{IR}},p_{\mathrm{SA}}\},S_{\mathrm{A}}^{\mathrm{T}}\right)-\gamma_3 R_{\mathrm{D}}^{\mathrm{T}}, \tag{18}$$

where $R_{\mathrm{D}}^{\mathrm{T}}$ represents the resources consumed by the defender, $\rho_{\mathrm{Traceback}}$ represents the probability of successful traceback (including identifying attack paths and attack intentions), and α_3 and γ_3 are weights.

Correspondingly, the attacker aims to evade traceback and extend the attack's duration by obfuscating the attack path and concealing behavioral characteristics. The attacker's strategy space is defined as $S_{\mathrm{A}}^{\mathrm{T}}=\{a_{\mathrm{PD}},a_{\mathrm{FO}}\}$. Here, a_{PD} represents path disturbance, which dynamically alters attack path nodes or communication chains to increase the difficulty of traceback; a_{FO} represents feature obfuscation, which utilizes obfuscation techniques to evade the feature recognition mechanisms applied by the defender.

The attacker seeks to maximize the attack's survival time while minimizing the resource consumption by applying path disturbance and feature obfuscation techniques. The payoff function is defined as

$$U_{\mathrm{A}}^{\mathrm{T}}=\delta_3\rho_{\mathrm{Evade}}\left(\theta_{\mathrm{A}},\{a_{\mathrm{PD}},a_{\mathrm{FO}}\},S_{\mathrm{D}}^{\mathrm{T}}\right)-\zeta_3 R_{\mathrm{A}}^{\mathrm{T}}, \tag{19}$$

where $R_{\mathrm{A}}^{\mathrm{T}}$ represents the resources consumed by the attacker, ρ_{Evade} represents the probability of successfully evading traceback, and δ_3 and ζ_3 are weights.

Guided by comprehensive path profiling, we have proposed a novel framework that integrates graph neural network-based anomaly process detection with provenance graph-driven attack path reconstruction (He ZX, 2024). We conducted comprehensive evaluations of our proposed method on the DARPA datasets, benchmarking against state-of-the-art approaches. In anomaly detection, our method achieved 95.45% and 96.24% F1-scores on DARPA CADETS and DARPA TRACE respectively, significantly outperforming ProvDetector (86.56%/88.88%) (Wang Q et al., 2020) and ATLAS (90.91%/91.11%) (Alsaheel et al., 2021). For attack path reconstruction, our solution demonstrated a superior coverage of 98.6% anomalous nodes, surpassing existing systems including SLEUTH (93.4%) (Hossain et al., 2017), HOLMES (95.8%) (Milajerdi et al., 2019), Nodone (86.0%) (Hassan et al., 2019), and ATLAS (84.8%). These results validate our method's enhanced detection accuracy and attack traceability capabilities in cybersecurity threat analysis.

5.4.4 Summary

In the SAPC model, comprehensive path profiling enhances attack traceback capabilities through three key mechanisms: attack path perception, attack intent reasoning, and traceback strategy optimization. By incorporating game-theoretic principles, it enables precise early-stage intervention in the attack chain while facilitating dynamic strategy adaptation for optimal resource allocation. It not only improves attack traceback efficiency at the technical level but also strengthens defensive strategic positioning, thereby contributing to the robustness of contemporary cybersecurity frameworks.

5.5 Dynamic deception

Dynamic countermeasures aim to force attackers to discontinue their attacks or reduce attack effectiveness by increasing the operational cost. These technologies work by depleting the attackers' resources, raising the complexity of the attacks, and reinforcing the defense by targeting attack sources and infrastructure and by implementing intelligent countermeasures. For the defender to select the most effective countermeasure, it is crucial to assess the attacker's behavior. In this context, game theory emerges as a potent tool for understanding and steering the deployment of countermeasures.

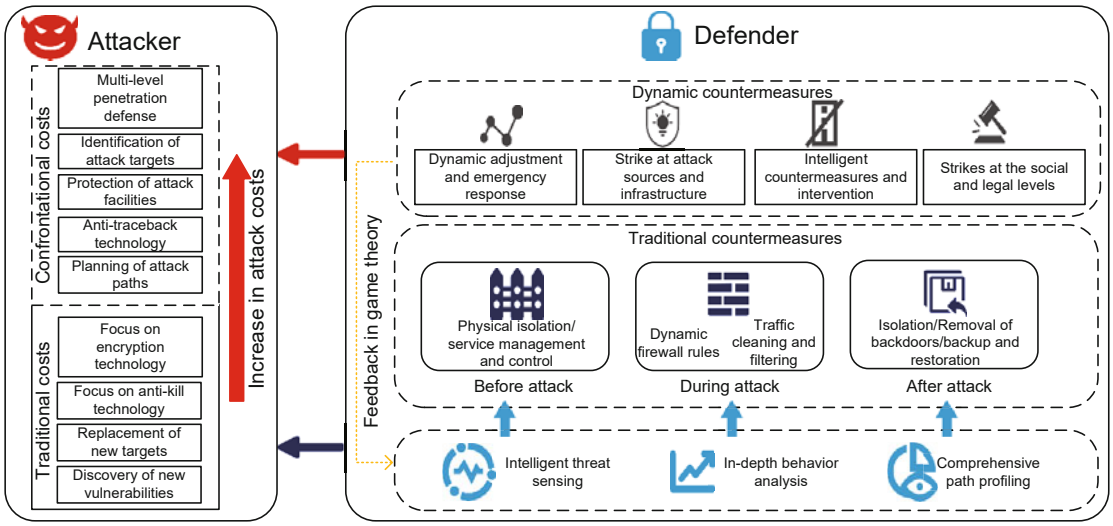

Fig. 6 Countermeasure technology based on active cybersecurity with game theory

5.5.1 Background

The lifecycle of a cyberattack, as shown in Fig. 6, is typically divided into three distinct stages: preattack, mid-attack, and postattack. Conventional cybersecurity methods primarily rely on passive defenses, which are less effective in proactively addressing and countering the actions of attackers during the attack phase.

During the preattack stage, attackers typically gather information and scan for vulnerabilities to identify weak points in their target, which helps them plan their attack routes. Traditional defense strategies aim to impede attackers by making information collection more difficult and reducing the exposure of public network assets through active measures such as physical isolation and service control, which ultimately increase attack costs. However, these strategies have limited effectiveness. In the mid-attack stage, attackers attempt to breach defenses, exploit vulnerabilities, or launch cyberattacks to fulfill their objectives. Traditional defense measures during this phase largely focus on real-time response and limiting the impact of attacks, such as adjusting firewall rules dynamically (Tudosi et al., 2023), blocking source IP addresses (Chinnasamy et al., 2023), and utilizing traffic cleaning and filtering devices (Yuan et al., 2024). Unfortunately, these responses are often slow and insufficient in addressing the increasingly complex and varied nature of attacks. In the postattack stage, attackers may have successfully gained persistent access to the system. Traditional countermeasures focus on reducing the attacker's control over the system and removing malicious components, such as isolating infected areas (Xiao JT et al., 2023), using scanning tools to detect and remove backdoor programs (Abelson et al., 2024), and restoring the system from backups (Hasan et al., 2023). However, these measures are reactive and do not prevent attackers from exploiting existing attack vectors to regain access.

5.5.2 Proposed technology

At its core, game theory focuses on analyzing strategic interactions between participants. In the realm of adversarial countermeasures for active network security, the defender and the attacker are treated as two opposing players, each with distinct strategies aimed at maximizing their benefits. The defender's objective is to escalate the cost of the attack, thereby forcing the attacker to either abandon the attack or reduce its scope due to the high resource investment required. On the other hand, the attacker strives to achieve their goals by minimizing costs and circumventing the defenses. Core technologies involved in adversarial countermeasures within active network security (Chen RD et al., 2019) include dynamic adjustments and emergency responses (Kaufhold et al., 2024), targeting attack sources and infrastructure (Chen S and Taw, 2023), intelligent

countermeasures and interventions (Pawlicki et al., 2023), and legal and social-level responses (Khater, 2023). These technologies work in tandem, complementing one another in effectively raising the cost of attack for the adversary.

5.5.3 Case study

In the game of adversarial countermeasures, the defender's objective is to increase the attacker's attack cost through a series of countermeasure actions and, if necessary, compel the attacker to halt the attack. The defender's strategy space is $S_{\mathrm{D}} = \{s_{\mathrm{rsp}}, s_{\mathrm{hack}}, s_{\mathrm{dec}}, \dots\}$, where each term refers to specific defensive actions. The term s_{rsp} represents the immediate response and dynamic adaptation, including vulnerability patching and resource reallocation, to mitigate the impact on the nodes compromised by the adversary. The strategy s_{hack} focuses on targeting and compromising the attacker's infrastructure. Finally, s_{dec} applies deceptive strategies, such as honeypots, to increase the adversary's operational costs and thereby further compel them to abort the attack. The payoff function is defined as

$$U_{\mathrm{D}}(\theta_{\mathrm{D}}, S_{\mathrm{D}}, S_{\mathrm{A}}, R_{\mathrm{D}}) = \sum_{i=1}^{n_{\mathrm{D}}} (\alpha_i \rho_i(S_{\mathrm{D}}^i, S_{\mathrm{A}}) - \beta_i R_{\mathrm{D}}^i), \tag{20}$$

where $\rho_i(S_{\mathrm{D}}^i, S_{\mathrm{A}})$ represents the probability of successfully countering the adversary and increasing the adversary's attack cost, and α_i and β_i are dimension parameters.

Correspondingly, when faced with the defender's countermeasure operations, the attacker will also implement a series of countermeasures. The attacker's strategy space is $S_{\mathrm{A}} = \{s_{\mathrm{repath}}, s_{\mathrm{rebuild}}, s_{\mathrm{prot}}, \dots\}$, where s_{repath} involves selecting a new attack path when the original path is compromised, s_{rebuild} refers to rebuilding the attack infrastructure or equipment if it is destroyed, and s_{prot} involves seeking political protection when faced with social strikes or legal countermeasures. The payoff function is defined as follows:

$$U_{\mathrm{A}}(\theta_{\mathrm{A}}, S_{\mathrm{A}}, S_{\mathrm{D}}, R_{\mathrm{A}}) = \sum_{i=1}^{n_{\mathrm{A}}} (\gamma_i \rho_i(S_{\mathrm{A}}, S_{\mathrm{D}}^i) - \delta_i R_{\mathrm{A}}^i), \tag{21}$$

where $\rho_i(S_{\mathrm{A}}, S_{\mathrm{D}}^i)$ represents the probability of successfully evading traceability, and γ_i and δ_i are dimension parameters.

Chen RD et al. (2019) presented a real-time detection and self-learning evolution technique based on zero-day vulnerability exploitation, aimed at continuously improving the detection and defense capabilities against attack samples. This technique effectively analyzes and blocks various APT-related traffic, significantly reducing false positives and false negatives. Based on this countermeasure strategy, we successfully identified and blocked nine attack tools used by the APT29 group. In the honeypot technology developed based on the idea of adversarial game, our model in a previous work (Niu et al., 2022) could more effectively counter attackers and gain the initiative. In our experiment, the attack deception effect increased by 11%.

5.5.4 Summary

The core technologies within the active network security framework are not standalone components; rather, they collaborate synergistically to form a multilayered, dynamically adaptive defense system. Internal defenses and external counterattacks complement one another, each playing a critical role in thwarting attacks. The internal defense mechanisms work by continuously adjusting the network environment, thereby increasing the complexity of attacks and depleting the attacker's resources. In parallel, external countermeasures directly target the attacker's infrastructure and attack sources, weakening their capabilities and hindering further progress. Together, these technologies create a closed-loop defense system. Should an attacker attempt to bypass a particular defense layer, other technologies within the system respond promptly, deploying dynamic countermeasures that raise the difficulty and cost of breaching the defense. This collaborative approach significantly enhances the overall resilience and effectiveness of the network's security posture.

5.6 Applications

Guided by the theoretical model of SAPC, incorporating core elements such as intelligent threat sensing, in-depth behavior analysis, comprehensive path profiling, and dynamic countermeasures, the model emphasizes proactive and adaptive defense

strategies. This approach is specifically designed to address cybersecurity challenges in complex attack scenarios, considering them from a multi-dimensional and full-lifecycle perspective. The model integrates critical technologies, including targeted vulnerability mining, adversarial traffic detection, game-theory-based traceability analysis, and active adversarial countermeasures. The ultimate goal is to increase the cost of attacks for adversaries, mitigate the damage resulting from successful breaches, and enhance the adversarial game capabilities of the entire information system.

To validate the feasibility and effectiveness of the proposed theoretical model, its practical application is demonstrated within a cloud computing environment, specifically using the cloud desktop management system based on security gateways and its associated security access control methods. The independently developed cloud computing servers, gateways, and desktop terminal products have achieved international leadership in key technical areas, including virtualization security access management and recovery scheduling performance. These solutions are widely deployed across critical national infrastructure sectors, such as electricity, healthcare, and education.

Notably, in significant national projects such as the West–East Power Transmission and the Sichuan–Xizang Power Grid Interconnection, these technologies have played a pivotal role in ensuring the stable operation of vital information systems, including power grid dispatching and production. They have also been instrumental in the rapid repair and recovery of power grids in disaster-stricken areas. Statistical evidence indicates that the deployment of these technologies has generated more than 1.5 billion yuan in direct economic benefits for related industries over the past eight years.

Moreover, during the global ransomware incident of 2017, the effectiveness of the SAPC model was clearly demonstrated. That year, ransomware attacked approximately 300 000 computers worldwide, encrypting vast amounts of critical data and posing a severe threat. Among the affected organizations was The Affiliated Hospital of Southwest Medical University in Sichuan Province, China, whose diagnosis and treatment system fell victim to the ransomware.

However, the hospital had already implemented a security virtualization system based on the SAPC model. This system, equipped with intelligent threat sensing technology, detected the ransomware intrusion in real time. It quickly traced the virus' transmission path and infection behavior through intelligent sensing and traffic monitoring.

Simultaneously, in the "in-depth analysis" phase, the system rapidly extracted virus samples and identified the malicious encryption behavior affecting the hospital's data. During the "comprehensive path profiling" phase, the system executed dynamic countermeasures to sever the malicious transmission channels and thereafter activated an emergency recovery mechanism.

Within three hours, the hospital's diagnosis and treatment system was fully restored, with no data loss or damage. This rapid recovery was possible due to the seamless coordination of the four phases of the framework: intelligent threat sensing, in-depth behavior analysis, comprehensive path profiling, and active response and countermeasures. This real-world application highlights the high efficiency and practicality of the SAPC model in mitigating threats and restoring system integrity.

Compared to traditional passive defense systems, the cases outlined above underscore the significant advantages of the SAPC model. In traditional systems, defense measures typically activate only after an attack has occurred, reacting to the breach rather than preventing it. In contrast, the SAPC model enhances attack prevention by deploying intelligent sensing, dynamic traceability, and active countermeasures in advance.

By incorporating game-theoretic analysis, the active model boosts the adaptability and decision-making capabilities of the defense system, allowing it to anticipate and counter threats more effectively. Unlike traditional defense methods, which often rely on virus databases and signature-matching technologies, the active security model is capable of defending against APTs, including unknown viruses and zero-day attacks.

Moreover, the active framework supports self-evolution and dynamic adjustment of defense strategies, overcoming the delays and limitations associated with traditional defense mechanisms and allowing the system to continuously adapt to evolving threats.

6 Challenges and future trends

As cybersecurity threats continue to evolve, traditional defense mechanisms demonstrate limitations in addressing advanced attacks. Active cybersecurity, as an emerging defensive paradigm, implements proactive threat identification and response through real-time monitoring, dynamic response, and intelligent prediction. However, the practical implementation of active cybersecurity frameworks encounters significant challenges across technical, organizational, and management levels. This section examines these challenges comprehensively, analyzes future development trends, and proposes relevant recommendations.

6.1 Challenges

The proliferation of devices and rapid application iteration has transformed security solutions from isolated network systems into complex, interconnected network ecosystems. Contemporary network environments encompass multimillion-node device clusters, heterogeneous application landscapes, and diverse network architectures, particularly evident in Internet of Things (IoT) infrastructures. This evolution of network systems necessitates a paradigm shift in security technologies, transitioning from passive, static defense mechanisms to active, dynamic security frameworks. Network security systems must demonstrate continuous adaptation capabilities within these complex environments to effectively address evolving application scenarios.

6.1.1 Incomplete collection of information elements in complex network environments

In complex network environments, cybersecurity faces a fundamental challenge of incomplete critical information acquisition. Complex networks consist of numerous nodes, intricate connections, massive traffic volumes, and variable traffic flows. The inherent flexibility and openness of network topological structures impede comprehensive monitoring of node status and identification of potential risk entry points. Attackers exploit multiple attack vectors, including covert channels, multihop connections, emerging application vulnerabilities, and protocol conversion vulnerabilities. Defenders, constrained by limited monitoring and analysis capabilities, must rely on restricted information sources such as partial network traffic data, known attack patterns, and vulnerability information. This information asymmetry impairs the defenders' ability to comprehensively assess attackers' intentions and capabilities, ultimately leading to suboptimal defensive decisions and delayed response mechanisms.

To address this issue, defenders require advanced game theory methods, such as signaling game theory, for continuous monitoring and analysis of network traffic at critical nodes and network links. This enables the detection of anomalous behaviors and attack patterns, which can then be interpreted as clues to decode the attackers' signals to help adapt strategies accordingly. Information sharing and cross-domain cooperation in complex network environments should be strengthened, and a data-sharing platform should be established to integrate data from industry, government, and international sources, enhancing the diversity of intelligence sources. Moreover, the integration of AI with automated defense mechanisms, including intelligent firewalls and behavioral analysis platforms, enables dynamic rule adaptation and context-aware user behavior analysis. This facilitates real-time analysis of attackers' signals based on complex network characteristics, enhancing response capabilities while minimizing manual intervention, thereby optimizing defensive effectiveness.

6.1.2 Complex calculations of massive data in complex network environments

In complex network environments, cybersecurity faces significant computational challenges in processing massive-scale data. Network attack–defense interactions generate extensive datasets from diverse sources, including device monitoring data, security system logs, and network traffic information. These data exhibit multi-dimensional characteristics, heterogeneous structures, and inherent noise, with volumes that exceed traditional data-processing capabilities. From a game theory perspective, defenders must derive attackers' strategies from these extensive datasets to formulate optimal responses. However, the inherent data complexity significantly impedes computational analysis. For example, in network traffic analysis, the volume and complexity of data interfere with accurate identification of attacker's behavior patterns. This computational complexity impairs defenders'

ability to extract attacker's intentions efficiently, resulting in delayed decision-making within the cybersecurity game. Consequently, defenders struggle to implement timely strategy adaptations in response to dynamic attackers' behaviors, thereby placing themselves in a passive position.

To address this challenge, defenders should implement an intelligent data processing system integrated with game theory principles. A distributed parallel computing architecture is essential, enabling the allocation of computing tasks through equilibrium strategies from game theory. This approach allows multiple computing units to collaborate in processing data, thereby enhancing processing speed. Moreover, game theory-based reinforcement learning algorithms can be applied to derive valuable insights from the data, continuously refining the identification of abnormal patterns and potential risks through iterative learning.

For instance, in detecting the propagation paths of malicious software, game-theoretic relationships between network nodes can guide the optimization of data search and analysis strategies. Simultaneously, cloud computing technology should be leveraged to scale computing resources elastically, ensuring efficient data processing during fluctuations in data volume. In addition, a robust data preprocessing mechanism should be established, applying game theory to optimize data cleaning and feature extraction, thereby reducing data complexity and enhancing quality. This integrated approach will provide defenders with a solid foundation for data-driven decision-making, strengthening their ability to respond to evolving and complex cybersecurity threats.

6.1.3 Complex emergency response coordination in complex network environments

Cybersecurity responsibility lies with various network operators, including network service providers, enterprise network management departments, and critical infrastructure operators. As network assets expand, attacks may originate from numerous unexpected nodes or routes. For example, in the global Internet architecture, if a small branch's network equipment is compromised, it can serve as a gateway for attackers to infiltrate the core network of a large enterprise or critical national infrastructure.

These entities face considerable challenges. On one hand, monitoring every detail within a network becomes increasingly difficult as the number of connection points and data traffic grow, exceeding the capacity of traditional monitoring tools. On the other hand, attackers often apply sophisticated encryption techniques, anonymous networks, and dynamic, adversarial strategies (Chakraborty et al., 2018), which makes it hard to quickly and accurately pinpoint the source and intent of attacks. This greatly diminishes the relevance and timeliness of emergency responses.

To address emergency response coordination challenges, the implementation of a centralized emergency response coordination and command center is essential. This center coordinates incident response resources, develops unified emergency strategies and action plans, defines stakeholder roles and responsibilities, and ensures efficient information transmission and sharing. Standardized emergency response processes and interface specifications should be developed to facilitate seamless collaboration among different organizations and systems within a unified framework. For instance, standardized formats and time requirements for reporting cybersecurity incidents, along with interface standards for linking various security products, should be established. Furthermore, cross-organizational and cross-industry emergency drills and training programs should be strengthened to foster trust and mutual understanding among all stakeholders, thereby enhancing their collective response capabilities. In addition, information technology should be utilized to create an emergency response information-sharing platform that consolidates data on emergency resources and security situations from all involved parties. This platform would enable real-time information exchange and resource allocation, ultimately improving the efficiency and effectiveness of emergency responses.

6.2 Future trends

In the future of cybersecurity, key trends will significantly shape the evolution of active security measures. This subsection examines these trends from three primary perspectives: technology, policy, and organization. These interconnected pillars are critical in determining how cybersecurity will adapt to address the increasingly complex and pervasive cyber threats.

With technological advancements in AI,

machine learning, and big data, active cybersecurity can evolve to become much more intelligent. First, using AI to improve threat detection and response is key. By integrating deep learning and behavioral analysis, abnormal behaviors can be detected with greater accuracy, and attack patterns can be forecasted, enabling more effective and automated defense measures. Second, the trend toward security automation continues to accelerate. Automated tools for threat detection, attack response, and patch management can drastically increase the speed of security interventions, reducing the impact of delays caused by manual processes. Furthermore, creating an adaptive security architecture is vital. A system capable of dynamically adjusting its protection strategies based on real-time network changes would significantly improve network security's flexibility and resilience.

Policy-level promotion is essential for advancing active cybersecurity. Strengthening cooperation at both the national and international levels is a critical approach to tackling global cybersecurity challenges. Governments, businesses, and organizations worldwide need to establish closer collaborative mechanisms, share cybersecurity intelligence, and collectively respond to transnational cyberattacks. Furthermore, the development and refinement of cybersecurity laws, regulations, and industry standards are crucial for enhancing security protection capabilities. By implementing robust regulations, governments can clearly define the responsibilities of all stakeholders, incentivize businesses to increase their cybersecurity investments, provide a framework for compliance, and foster technological innovation. To ensure the effectiveness of security protections, establishing a comprehensive security assessment and certification system is vital. Such a system would validate the effectiveness of various security products and solutions, offering enterprises a trust guarantee that their cybersecurity measures are both effective and reliable.

In terms of organizational development, strengthening the construction of a security culture and cultivating cybersecurity talent are essential components. To enhance cybersecurity culture, it is crucial to first improve employees' security awareness through training. This ensures that all staff members adhere to security regulations and are equipped to identify and respond to potential cybersecurity threats. As cybersecurity threats become increasingly sophisticated, the shortage of skilled professionals has emerged as a significant bottleneck in enhancing security capabilities. Therefore, collaboration among governments, enterprises, and educational institutions is particularly important. By jointly cultivating cybersecurity professionals, sufficient technical support can be provided to the industry, ensuring the long-term growth and resilience of the cybersecurity sector.

Through the multifaceted advancement of technology, policy, and organizational frameworks, active cybersecurity will evolve toward a more intelligent, flexible, and efficient development trajectory. This will empower all stakeholders to effectively address the increasingly complex cybersecurity threats they face.

7 Conclusions

The implementation of active cybersecurity marks a paradigm shift in the cybersecurity landscape, transitioning from traditional passive defense mechanisms to proactive prediction and prevention strategies, thereby significantly enhancing system security posture. By establishing a proactive and adaptive network defense architecture, active cybersecurity enables the early detection and mitigation of emerging threats. This is the core function of the intelligent threat sensing element, which leverages real-time collection and analysis of multi-source intelligence to actively identify threats and bridge information gaps. The in-depth behavior analysis element uncovers the attacker's behavioral patterns and tactical inclinations through a comprehensive analysis of multi-dimensional data, providing vital decision support for crafting precise defense strategies. The comprehensive path profiling element reveals the attacker's true intent and behavioral trajectory by dynamically tracing attack pathways and identifying critical nodes. This process lays the foundation for dynamic countermeasure strategies, and further strengthens the defensive response. Additionally, the dynamic countermeasure element enhances the defender's strategic initiative by continuously adapting defense strategies in response to the attacker's actions. By countering the attacker's resources and increasing the cost of their actions, it significantly reduces the threat posed by the attack, diminishing

its overall impact.

This robust defense strategy enhances the ability to identify and mitigate unforeseen threats while facilitating the accurate prediction of and effective response to adversarial actions, particularly in environments marked by information asymmetry. This is achieved through the application of dynamic game optimization techniques. The implementation of active cybersecurity marks a significant evolution in cybersecurity technologies, moving from traditional reactive defense mechanisms to a more intelligent and proactive approach centered on prediction and prevention. This shift offers substantial support for protecting the critical infrastructure, ensuring social stability, and mitigating risks in business operations. As technology continues to advance, active cybersecurity is poised to become the cornerstone in addressing cyber threats in the digital age. It effectively counters increasingly sophisticated, covert, and rapidly evolving attacks from adversaries through precise threat forecast, strategic planning optimization, and a comprehensive and multilayered defense framework. The realization of active cybersecurity also introduces innovative theoretical models and practical tools for the global governance of cyberspace security, ensuring the integrity and stability of cyberspace.

Contributors

Xiaosong ZHANG conceived the project. Yukun ZHU and Junpeng HE drafted the paper. Ran YAN, Shiping HUANG, and Fenghua XU helped organize the paper. Xiong LI, Weina NIU, and Yongzhao ZHANG revised and finalized the paper.

Conflict of interest

All the authors declare that they have no conflict of interest.

References

Abelson H, Anderson R, Bellovin SM, et al., 2024. Bugs in our pockets: the risks of client-side scanning. *J Cybersecur*, 10(1):tyad020. https://doi.org/10.1093/cybsec/tyad020

Alsaheel A, Nan YH, Ma SQ, et al., 2021. ATLAS: a sequence-based learning approach for attack investigation. Proc 30th USENIX Security Symp, p.3005-3022.

Arjunan T, 2024. Real-time detection of network traffic anomalies in big data environments using deep learning models. *Int J Res Appl Sci Eng Technol*, 12(3):844-850. https://doi.org/10.22214/ijraset.2024.58946

Bocovich C, Breault A, Fifield D, et al., 2024. Snowflake, a censorship circumvention system using temporary WebRTC proxies. Proc 33rd USENIX Conf on Security Symp, Article 148.

Cai GL, Wang BS, Hu W, et al., 2016. Moving target defense: state of the art and characteristics. *Front Inform Technol Electron Eng*, 17(11):1122-1153. https://doi.org/10.1631/FITEE.1601321

Chakraborty A, Alam M, Dey V, et al., 2018. Adversarial attacks and defences: a survey. https://arxiv.org/abs/1810.00069

Chao DC, Xu DW, Gao G, et al., 2024. A systematic survey on security in anonymity networks: vulnerabilities, attacks, defenses, and formalization. *IEEE Commun Surv Tutor*, 26(3):1775-1829. https://doi.org/10.1109/COMST.2024.3350006

Chen RD, Zhang XS, Niu WN, et al., 2019. A research on architecture of APT attack detection and countering technology. *J Univ Electron Sci Technol China*, 48(6):870-879 (in Chinese). https://doi.org/10.3969/j.issn.1001-0548.2019.06.011

Chen S, Taw J, 2023. Conventional retaliation and cyber attacks. *Cyber Def Rev*, 8(1):67-86.

Chinnasamy P, Devika S, Balaji V, et al., 2023. BDDoS: blocking distributed denial of service flooding attacks with dynamic path detectors. Proc Int Conf on Computer Communication and Informatics, p.1-5. https://doi.org/10.1109/ICCCI56745.2023.10128499

Crandall JR, Su ZD, Wu SF, et al., 2005. On deriving unknown vulnerabilities from zero-day polymorphic and metamorphic worm exploits. Proc 12th ACM Conf on Computer and Communications Security, p.235-248. https://doi.org/10.1145/1102120.1102152

Ding KY, Liu XL, Niu WN, et al., 2021. A low-query black-box adversarial attack based on transferability. *Knowl-Based Syst*, 226:107102. https://doi.org/10.1016/j.knosys.2021.107102

Fang BX, Jia Y, Li AP, et al., 2024. SARPPR: reconstructing cyberspace security defense model. *J Cybersecur*, 2(1):2-12 (in Chinese). https://doi.org/10.20172/j.issn.2097-3136.240101

Fowler C, Goffin M, Hill B, et al., 2020. An Introduction to MITRE Shield. The MITRE Corporation, USA.

Gao Y, 2012. Design of a security monitoring system for power information intranet based on the PDR2A model. *J Fujian Comput*, 28(7):137-138 (in Chinese). https://doi.org/10.3969/j.issn.1673-2782.2012.07.063

Goodfellow IJ, Shlens J, Szegedy C, 2014. Explaining and harnessing adversarial examples. Proc 3rd Int Conf on Learning Representations.

Han WJ, Xue JF, Wang Y, et al., 2021. APTMalInsight: identify and cognize APT malware based on system call information and ontology knowledge framework. *Inform Sci*, 546:633-664. https://doi.org/10.1016/j.ins.2020.08.095

Hand R, Ton M, Keller E, 2013. Active security. Proc 12th ACM Workshop on Hot Topics in Networks, Article 17. https://doi.org/10.1145/2535771.2535794

Harsanyi JC, 1967. Games with incomplete information played by "Bayesian" players, I–III part I. the basic model. *Manag Sci*, 14(3):159-182. https://doi.org/10.1287/mnsc.14.3.159

Hasan MZ, Sarwar N, Alam I, et al., 2023. Data recovery and backup management: a cloud computing impact. Proc IEEE Int Conf on Emerging Trends in Engineering, Sciences and Technology, p.1-6. https://doi.org/10.1109/ICEST56843.2023.10138852

Hassan WU, Guo SJ, Li D, et al., 2019. NoDoze: combatting threat alert fatigue with automated provenance triage. Proc 26th Annual Network and Distributed System Security Symp, p.487-504. https://doi.org/10.14722/ndss.2019.23349

He JP, Luo L, Xiao K, et al., 2022. Generate qualified adversarial attacks and foster enhanced models based on generative adversarial networks. *Intell Data Anal*, 26(5):1359-1377. https://doi.org/10.3233/IDA-216134

He K, Kim DD, Asghar MR, 2023. Adversarial machine learning for network intrusion detection systems: a comprehensive survey. *IEEE Commun Surv Tutor*, 25(1):538-566. https://doi.org/10.1109/COMST.2022.3233793

He ZX, 2024. Research on Attack Scenario Reconstruction Based on Heterogeneous Graph Attention Network. MS Thesis, University of Electronic Science and Technology of China, Chengdu, China (in Chinese). https://doi.org/10.27005/d.cnki.gdzku.2024.005404

Heartfield R, Loukas G, 2016. A taxonomy of attacks and a survey of defence mechanisms for semantic social engineering attacks. *ACM Comput Surv*, 48(3):1-39. https://doi.org/10.1145/2835375

Hossain N, Milajerdi SM, Wang JN, et al., 2017. SLEUTH: real-time attack scenario reconstruction from COTS audit data. Proc 26th USENIX Conf on Security Symp, p.487-504.

Hu HC, Sui JQ, Zhang S, et al., 2024. Proactive defense technology in cyber security: strategies, methods and challenges. *Comput Sci*, 51(S2):829-831 (in Chinese). https://doi.org/10.11896/jsjkx.231100132

Huang LN, Zhu QY, 2020. A dynamic games approach to proactive defense strategies against advanced persistent threats in cyber-physical systems. *Comput Secur*, 89:101660. https://doi.org/10.1016/j.cose.2019.101660

Jia Y, Qi YL, Shang HJ, et al., 2018. A practical approach to constructing a knowledge graph for cybersecurity. *Engineering*, 4(1):53-60. https://doi.org/10.1016/j.eng.2018.01.004

Jiang JC, Ma HT, Ren DE, et al., 2000. A survey of intrusion detection research on network security. *J Softw*, 11(11):1460-1466 (in Chinese). https://doi.org/10.13328/j.cnki.jos.2000.11.005

Jiang JG, Wang JZ, Kong B, et al., 2018. On the survey of network attack source traceback. *J Cyber Secur*, 3(1):111-131 (in Chinese). https://doi.org/10.19363/j.cnki.cn10-1380/tn.2018.01.008

Jiang X, 2020. Research on dynamic host security protection platform based on EDR and CARTA model. *Netw Secur Technol Appl*, (9):47-48 (in Chinese). https://doi.org/10.3969/j.issn.1009-6833.2020.09.032

Kaufhold MA, Riebe T, Bayer M, et al., 2024. 'We do not have the capacity to monitor all media': a design case study on cyber situational awareness in computer emergency response teams. Proc CHI Conf on Human Factors in Computing Systems, Article 580. https://doi.org/10.1145/3613904.3642368

Kaur R, Gabrijelčič D, Klobučar T, 2023. Artificial intelligence for cybersecurity: literature review and future research directions. *Inform Fus*, 97:101804. https://doi.org/10.1016/j.inffus.2023.101804

Khater MH, 2023. International perspective on securing cyberspace against terrorist acts. *Int J Sociotechnol Knowl Dev*, 15(1):1-11. https://doi.org/10.4018/IJSKD.318706

Kheddar H, Himeur Y, Awad AI, 2023. Deep transfer learning for intrusion detection in industrial control networks: a comprehensive review. *J Netw Comput Appl*, 220:103760. https://doi.org/10.1016/j.jnca.2023.103760

Kheddar H, Hemis M, Himeur Y, et al., 2024. Deep learning for steganalysis of diverse data types: a review of methods, taxonomy, challenges and future directions. *Neurocomputing*, 581:127528. https://doi.org/10.1016/j.neucom.2024.127528

Kim T, Park N, Hong J, et al., 2022. Phishing URL detection: a network-based approach robust to evasion. Proc ACM SIGSAC Conf on Computer and Communications Security, p.1769-1782. https://doi.org/10.1145/3548606.3560615

Kröse BJA, 1995. Learning from delayed rewards. *Robot Auton Syst*, 15(4):233-235. https://doi.org/10.1016/0921-8890(95)00026-C

Kumar R, Agrawal N, 2023. Analysis of multi-dimensional industrial IoT (IIoT) data in edge–fog–cloud based architectural frameworks: a survey on current state and research challenges. *J Ind Inform Integr*, 35:100504. https://doi.org/10.1016/j.jii.2023.100504

Li DP, Aung Z, Williams J, et al., 2014. P2DR: privacy-preserving demand response system in smart grids. Proc Int Conf on Computing, Networking and Communications, p.41-47. https://doi.org/10.1109/ICCNC.2014.6785302

Li HH, Zhang SG, Song H, et al., 2021. Robust malicious encrypted traffic detection based with multiple features. *J Cyber Secur*, 6(2):129-142 (in Chinese). https://doi.org/10.19363/J.cnki.cn10-1380/tn.2021.03.09

Li PY, Li X, Chen JJ, et al., 2022. Adversarial sample generation for evading botnet traffic detection. *Comput Eng Appl*, 58(4):126-133 (in Chinese). https://doi.org/10.3778/j.issn.1002-8331.2008-0298

Liang WT, Ling X, Wu JZ, et al., 2023. A needle is an outlier in a haystack: hunting malicious PyPI packages with code clustering. Proc 38th IEEE/ACM Int Conf on Automated Software Engineering, p.307-318. https://doi.org/10.1109/ASE56229.2023.00085

Lin ZL, Shi Y, Xue Z, 2022. IDSGAN: generative adversarial networks for attack generation against intrusion detection. Proc 26th Pacific-Asia Conf on Advances in Knowledge Discovery and Data Mining, p.79-91. https://doi.org/10.1007/978-3-031-05981-0_7

Liu QX, Wang JN, Yin J, et al., 2021. Application of adversarial machine learning in network intrusion detection. *J Commun*, 42(11):1-12 (in Chinese). https://doi.org/10.11959/j.issn.1000-436x.2021193

Liu XL, Wan K, Ding YF, et al., 2020. Weighted-sampling audio adversarial example attack. Proc 34th AAAI Conf on Artificial Intelligence, p.4908-4915. https://doi.org/10.1609/aaai.v34i04.5928

Manshaei MH, Zhu QY, Alpcan T, et al., 2013. Game theory meets network security and privacy. *ACM Comput Surv*, 45(3):25. https://doi.org/10.1145/2480741.2480742

Milajerdi SM, Gjomemo R, Eshete B, et al., 2019. HOLMES: real-time APT detection through correlation of suspicious information flows. Proc IEEE Symp on Security and Privacy, p.1137-1152. https://doi.org/10.1109/SP.2019.00026

Nash JF, 2002. Non-cooperative games. In: Bridel P (Ed.), The Foundations of Price Theory, Vol 4. Routledge, London, UK, p.329-340. https://doi.org/10.4324/9781003547983

Niu WN, Zhou J, Zhao YB, et al., 2022. Uncovering APT malware traffic using deep learning combined with time sequence and association analysis. *Comput Secur*, 120:102809. https://doi.org/10.1016/j.cose.2022.102809

Oh SE, Yang TJ, Mathews N, et al., 2022. DeepCoFFEA: improved flow correlation attacks on Tor via metric learning and amplification. Proc IEEE Symp on Security and Privacy, p.1915-1932. https://doi.org/10.1109/SP46214.2022.9833801

Pawlicki M, Pawlicka A, Kozik R, et al., 2023. The survey and meta-analysis of the attacks, transgressions, countermeasures and security aspects common to the cloud, edge and IoT. *Neurocomputing*, 551:126533. https://doi.org/10.1016/j.neucom.2023.126533

Pouyanfar S, Sadiq S, Yan YL, et al., 2019. A survey on deep learning: algorithms, techniques, and applications. *ACM Comput Surv*, 51(5):92. https://doi.org/10.1145/3234150

Rajapaksha S, Kalutarage H, Al-Kadri MO, et al., 2023. AI-based intrusion detection systems for in-vehicle networks: a survey. *ACM Comput Surv*, 55(11):237. https://doi.org/10.1145/3570954

Sabnis S, Verbruggen M, Hickey J, et al., 2012. Intrinsically secure next-generation networks. *Bell Labs Techn J*, 17(3):17-36. https://doi.org/10.1002/bltj.21556

Satvat K, Gjomemo R, Venkatakrishnan VN, 2021. Extractor: extracting attack behavior from threat reports. Proc IEEE European Symp on Security and Privacy, p.598-615. https://doi.org/10.1109/EuroSP51992.2021.00046

Schwartau W, 1998. Time-based security explained: provable security models and formulas for the practitioner and vendor. *Comput Secur*, 17(8):693-714. https://doi.org/10.1016/S0167-4048(98)80100-4

Shi C, Peng JH, Zhu SY, et al., 2024. From passive defense to proactive defence: strategies and technologies. Proc 1st Int Conf on Artificial Intelligence Security and Privacy, p.190-205. https://doi.org/10.1007/978-981-99-9785-5_14

Strom BE, Applebaum A, Miller DP, et al., 2020. MITRE ATT&CK®: Design and Philosophy. Project No. 10AOH08A-JC, The MITRE Corporation, McLean, USA.

Sun C, Hu H, Yang YJ, et al., 2022. Prediction method of 0day attack path based on cyber defense knowledge graph. *Chin J Netw Inform Secur*, 8(1):151-166 (in Chinese). https://doi.org/10.11959/j.issn.2096-109x.2021101

Sun S, Zhang L, Hu CH, et al., 2023. Cyberspace security models and systematic development from multiple perspectives. *Strat Study CAE*, 25(6):116-125 (in Chinese). https://doi.org/10.15302/J-SSCAE-2023.06.009

Sutton M, Greene A, Amini P, 2007. Fuzzing: Brute Force Vulnerability Discovery. Addison-Wesley Professional, Boston, USA.

Tan JL, Jin H, Zhang HQ, et al., 2023. A survey: when moving target defense meets game theory. *Comput Sci Rev*, 48:100544. https://doi.org/10.1016/j.cosrev.2023.100544

Tirpak JA, 2000. Find, fix, track, target, engage, assess. *Air Force Mag*, 83(7):24-29.

Tudosi AD, Graur A, Balan DG, et al., 2023. Design and implementation of an automated dynamic rule system for distributed firewalls. *Adv Electr Comput Eng*, 23(3):29-38. https://doi.org/10.4316/AECE.2023.03004

Wang D, Zhang XS, Chen T, 2020. Research on discovering memory corruption vulnerabilities for embedded CGIs. *J Univ Electron Sci Technol China*, 49(5):745-750 (in Chinese). https://doi.org/10.12178/1001-0548.2019233

Wang J, Huang ZS, Liu HL, et al., 2023. DefectHunter: a novel LLM-driven boosted-conformer-based code vulnerability detection mechanism. https://arxiv.org/abs/2309.15324

Wang Q, Hassan WU, Li D, et al., 2020. You are what you do: hunting stealthy malware via data provenance analysis. Proc 27th Annual Network and Distributed System Security Symp, p.1-17. https://doi.org/10.14722/ndss.2020.24167

Wang XD, Sun K, Batcheller A, et al., 2019. Detecting "0-day" vulnerability: an empirical study of secret security patch in OSS. Proc 49th Annual IEEE/IFIP Int Conf on Dependable Systems and Networks, p.485-492. https://doi.org/10.1109/DSN.2019.00056

Wei CK, Meng WL, Zhang ZK, et al., 2024. LMSanitator: defending prompt-tuning against task-agnostic backdoors. Proc 31st Annual Network and Distributed System Security Symp, p.1-18. https://doi.org/10.14722/ndss.2024.23238

Willbold J, Schloegel M, Vögele M, et al., 2023. Space Odyssey: an experimental software security analysis of satellites. Proc IEEE Symp on Security and Privacy, p.1-19. https://doi.org/10.1109/SP46215.2023.10351029

Wu JX, 2016. Research on cyber mimic defense. *J Cyber Secur*, 1(4):1-10 (in Chinese). https://doi.org/10.19363/j.cnki.cn10-1380/tn.2016.04.001

Xia Y, Lang RL, Dai GZ, 2001. Research on detect technology of intrusion detection system. *Comput Eng Appl*, 37(24):32-34, 118 (in Chinese). https://doi.org/10.3321/j.issn:1002-8331.2001.24.013

Xiao CW, Li B, Zhu JY, et al., 2018. Generating adversarial examples with adversarial networks. Proc 27th Int Joint Conf on Artificial Intelligence, p.3905-3911.

Xiao JT, Yang NZ, Shen WB, et al., 2023. Attacks are forwarded: breaking the isolation of MicroVM-based containers through operation forwarding. Proc 32nd USENIX Conf on Security Symp, Article 421.

Xiong CL, Zhu TT, Dong WH, et al., 2022. Conan: a practical real-time APT detection system with high accuracy and efficiency. *IEEE Trans Depend Secur Comput*, 19(1):551-565. https://doi.org/10.1109/TDSC.2020.2971484

Xu K, Tang M, Wang QC, et al., 2024. Exploitation of security vulnerability on retirement. Proc IEEE Int Symp on High-Performance Computer Architecture, p.1-14. https://doi.org/10.1109/HPCA57654.2024.00012

Xu XZ, Zeng X, Niu YF, 2024. Research on risk assessment and countermeasures for university network security based on the APPDRR model. *Netw Secur Technol Appl*, (4):89-93 (in Chinese). https://doi.org/10.3969/j.issn.1009-6833.2024.04.029

Yan Q, Wang MD, Huang WY, et al., 2019. Automatically synthesizing DoS attack traces using generative adversarial networks. *Int J Mach Learn Cyber*, 10(12):3387-3396. https://doi.org/10.1007/s13042-019-00925-6

Yang TF, Qiao YS, Lee B, 2024. Towards trustworthy cybersecurity operations using Bayesian deep learning to improve uncertainty quantification of anomaly detection. *Comput Secur*, 144:103909. https://doi.org/10.1016/j.cose.2024.103909

Yang Y, Sun L, Zhang CC, et al., 2024. Research on dynamic data security protection model based on Petri nets. Proc Int Conf on Machine Intelligence and Digital Applications, p.155-161. https://doi.org/10.1145/3662739.367085

Yao CJ, 2010. Applications of WPDRRC information security model in multi-level security protection. *Study Opt Commun*, (5):27-29 (in Chinese). https://doi.org/10.3969/j.issn.1005-8788.2010.05.009

Yuan QJ, Zhu YF, Xiong G, et al., 2024. ULDC: unsupervised learning-based data cleaning for malicious traffic with high noise. *Comput J*, 67(3):976-987. https://doi.org/10.1093/comjnl/bxad036

Zeng J, Chua ZL, Chen YF, et al., 2021. WATSON: abstracting behaviors from audit logs via aggregation of contextual semantics. Proc 28th Annual Network and Distributed System Security Symp, p.1-18. https://doi.org/10.14722/ndss.2021.24549

Zengy J, Wang X, Liu JH, et al., 2022. ShadeWatcher: recommendation-guided cyber threat analysis using system audit records. Proc IEEE Symp on Security and Privacy, p.489-506. https://doi.org/10.1109/SP46214.2022.9833669

Zhang B, Zhang ZY, Cheng LJ, et al., 2023. Topological characterization based on network traffic and DR attacking. *Commun Technol*, 56(4):494-501 (in Chinese). https://doi.org/10.3969/j.issn.1002-0802.2023.04.014

Zhang LD, Hemberg E, 2019. Investigating algorithms for finding Nash equilibria in cyber security problems. Proc Genetic and Evolutionary Computation Conf Companion, p.1659-1667. https://doi.org/10.1145/3319619.3326851

Zhang X, Shang JT, Liu ZJ, 2023. Research on network security protection system of scientific research institutes based on IPDRR model. *Netw Secur Technol Appl*, 12:127-129 (in Chinese).

Zhang Y, Hong JI, Cranor LF, 2007. CANTINA: a content-based approach to detecting phishing web sites. Proc 16th Int Conf on World Wide Web, p.639-648. https://doi.org/10.1145/1242572.1242659

Zhao J, Yan QB, Liu XD, et al., 2020. Cyber threat intelligence modeling based on heterogeneous graph convolutional network. Proc 23rd Int Symp on Research in Attacks, Intrusions and Defenses, p.241-256.

Zhou J, Ke P, Qiu XP, et al., 2024. ChatGPT: potential, prospects, and limitations. *Front Inform Technol Electron Eng*, 25(1):6-11. https://doi.org/10.1631/FITEE.2300089

Zhuo ZL, Zhang Y, Zhang ZL, et al., 2018. Website fingerprinting attack on anonymity networks based on profile hidden Markov model. *IEEE Trans Inform Forens Secur*, 13(5):1081-1095. https://doi.org/10.1109/TIFS.2017.2762825

Appendix: Explanation of core terms

Table A1 Explanation of core terms

No.	Term	Description
1	Active cybersecurity	A comprehensive security defense framework that combines active and passive strategies and technologies to effectively address network threats and attacks
2	SAPC model	The conceptual model implemented in this paper. Based on the game theory, this model iteratively optimizes the defense strategies through four key elements: intelligent threat sensing, in-depth behavior analysis, comprehensive path profiling, and dynamic countermeasures
3	Intelligent threat sensing	One key element in the SAPC model, which implements effective defensive measures prior to an attack, achieved by early identification and warning of potential threats
4	In-depth behavior analysis	One key element in the SAPC model, which identifies potential attack indicators, achieved by extracting critical threat information from its multi-dimensional characteristics
5	Comprehensive path profiling	One key element in the SAPC model, which reveals the attack path and the identity of the attacker by analyzing the behavioral trajectory, key nodes, and potential threats
6	Dynamic countermeasures	One key element in the SAPC model, which achieves the dynamic adjustment and enhancement of defense strategies, thereby raising the cost of the attacking tactics
7	Intrinsic security	An intrinsic mechanism of self-adaptation, autonomous defense, and dynamic optimization. Within this mechanism, the defense system is equipped with a security framework and technical architecture capable of proactively perceiving threats, rapidly responding to attacks, and continuously enhancing defensive capabilities
8	Mimic security	An internal mechanism that dynamically and pseudo-randomly selects and executes diverse hardware variants and their corresponding software configurations under both active and passive trigger conditions. This mechanism introduces significant uncertainty into the hardware execution environment and software operational states, thereby increasing the complexity and difficulty of establishing an attack chain based on vulnerabilities or backdoors
9	APT (advanced persistent threat)	A highly sophisticated, covert, and persistent cyberattack orchestrated by an organized entity, typically a state-sponsored group or a well-structured cybercriminal organization, aimed at exfiltrating sensitive information or disrupting critical infrastructure systems
10	F2T2EA (find, fix, track, target, engage, and assess)	A military kill chain attack framework encompassing six sequential phases: the search phase, which focuses on identifying the target; the fixation phase, dedicated to confirming the precise location of the target; the tracking phase, involving continuous monitoring of the target's movements; the targeting phase, where the appropriate weapon or asset is selected for deployment; the attack phase, which entails executing the attack using the chosen weapon; the assessment phase, aimed at evaluating the effectiveness and impact of the attack
11	ATT&CK (adversarial tactics, techniques, and common knowledge)	An attack matrix framework derived from empirically observed network attack characteristics. This model categorizes attack activities into stages such as initial access, execution, persistence, privilege escalation, defense evasion, credential access, discovery, lateral movement, collection, command and control, exfiltration, and impact, while providing detailed methods for implementing each stage of the attack
12	OSINT (open source intelligence)	An intelligence collection method applied by the United States Central Intelligence Agency to identify and extract valuable information from diverse publicly available sources
13	Tor (the onion router)	A privacy protection technology that conceals a user identity and location by transmitting data anonymously through multiple layers of encryption and distributed network relay nodes, where the data undergo sequential encryption and decryption, akin to peeling layers of an onion
14	C&C (command and control) server	The central server responsible for managing and controlling the botnet to communicate with individual zombie nodes and orchestrate the attack activities
15	Vulnerability mining	The process of actively identifying potential security vulnerabilities in systems, software, or networks using technical approaches or analytical methods to mitigate or address potential security risks
16	AE (adversarial example)	An instance with deliberately minor feature perturbations designed to mislead a machine learning model into making an incorrect prediction
17	Fuzzing	A technique for identifying software vulnerabilities by supplying unexpected inputs to a target system and observing anomalous responses
18	ISP (Internet service provider)	A telecommunication provider that offers Internet connectivity, information services, and value-added services to a broad spectrum of users. Within the Internet application service industry, the ISP functions as a content collector, producer, and service provider
19	IDS (intrusion detection system)	A security approach employed to observe and detect anomalous activities and security events within networks and computer systems