Industry Science and Engineering Vol. 3 No. 8, 2026

Research on Network-Attack Behavior Capture and Traceability Technology Based on Automatic-Capture Traceable Honeypot Platform

Zongchuan Feng[1,2], Linrong Yan[1*], Jie Qiu[1]
[1]*School of Artificial Intelligence, Yulin Normal University, Yulin, Guangxi, China*
[2]*Yulin Normal University Education Technology Center, Yulin, Guangxi, China*
**Corresponding Author*

Abstract: Cyberattacks are evolving toward industrialization, diversification and concealment. Complex multi-stage attack chains, incomplete retention of intrusion behaviors and difficulties in forensic traceability have become urgent challenges in cybersecurity defense. Traditional static testbeds can only reproduce standardized vulnerability-exploitation procedures; they struggle to fully capture full-life-cycle attack behaviors and restore real-world attack chains, which results in incomplete forensic evidence chains. To address these limitations, this paper proposes a technical scheme based on a high-interaction automatic-capture traceable honeypot platform for attack deception, multi-dimensional behavior collection and automatic link traceability. A three-layer technical architecture (bottom simulation-support layer, middle data-processing layer and top analysis-output layer) is constructed to realize business-service simulation, full-volume attack-behavior capture, automatic attack-link traceability and isolated experimental-domain control. A closed-loop workflow of "attack deception-raw-data acquisition-multi-source data correlation-trace-link generation-forensic-evidence export" is designed. The platform records attack-source information, operation commands, tool parameters, network traffic and time-series trajectories comprehensively, and generates visual attack-trace graphs automatically. Comparative experiments are conducted to validate data-capture completeness and trace-link-restoration performance. Experimental results demonstrate that the proposed scheme can preserve full-life-cycle attack behaviors including failed probing attempts, reconstruct complete attack chains, and mitigate technical drawbacks such as fragmented attack records and manually-assembled trace evidence in conventional simulation environments. This research provides technical support and experimental paradigms for network-intrusion behavior analysis, attack traceability and cyber-threat perception.

Keywords: High-Interaction Honeypot; Automatic Capture; Attack Traceability; Behavior Collection; Threat Perception; Digital Forensics

1. Introduction

With the rapid development of digital informatization, network business architectures across industries are becoming increasingly complicated. Driven by underground cybercrime economies, modern network attacks exhibit industrialization, tool-oriented implementation and strong concealment. Adversaries commonly adopt multi-phased intrusion, multi-hop pivoting and tool obfuscation tactics, which bring great challenges to attack-behavior reconstruction, threat traceability and digital-evidence preservation [1].

In cyber-threat research communities, attack reproduction is mostly implemented via static testbeds. Static testbeds can reproduce exploitation processes of known vulnerabilities, yet their scenarios are pre-configured. They only execute predefined test cases and fail to capture diverse and stochastic adversarial behaviors. Intermediate probing trials and unsuccessful intrusion attempts are frequently discarded, leading to fragmented attack records and discontinuous forensic evidence chains. Moreover, conventional simulation environments lack network-isolation mechanisms. Attack-simulation experiments carry risks of traffic leakage toward external

networks and unauthorized cross-domain access, restricting large-scale research on attack-data collection and traceability techniques [2].

As an active deception-based security-perception technology, high-interaction honeypots simulate real business services to lure intrusions. They possess inherent advantages in attack-behavior collection, intrusion-trajectory logging and environment isolation, making them vital experimental carriers for cyber-attack analysis and traceability research. Nevertheless, most existing honeypot systems focus merely on network-traffic collection [3]. Deficiencies remain in automatic concatenation of multi-stage attack links, time-axis trajectory restoration and full-persistence of application-layer operation sequences.

This paper designs an automatic-capture traceable honeypot technical solution. It focuses on key technologies including full-life-cycle attack-data acquisition, multi-dimensional information correlation, automatic attack-chain reconstruction and isolated simulation-domain management [4]. A complete closed-loop for attack-data processing is established. By simulating diverse business services, multi-source data such as attacker IPs, operation sequences, tool parameters and time-series logs are collected, and visual trace graphs are automatically generated to restore end-to-end intrusion procedures. Comparative experiments are performed to evaluate data-capture integrity and attack-chain-reconstruction capability. The outcomes offer experimental support for cyber-threat analysis and attack-forensic research.

Recent journal studies examine adaptive interaction in multi-step attack scenarios, survey cyber deception techniques, investigate honeynet traffic isolation through network slicing, and assess interaction levels in PLC honeypots. These studies provide complementary context for the platform's interaction, capture and isolation design.

2. Technical Limitations of Conventional Attack-Reproduction and Traceability Approaches

Based on state-of-the-art research on threat reproduction and attack traceability, static testbeds and traditional low-interaction honeypots have evident technical bottlenecks in attack-behavior logging and trace restoration, mainly reflected in scenario simulation, data acquisition, multi-source data correlation and experimental-environment control.

2.1 Insufficient Fidelity of Business-Scenario Simulation

Static testbeds adopt pre-defined vulnerability configurations and simple network topologies, which differ substantially from real-world Web sites, database clusters and enterprise-grade server deployments [5]. Experiments are constrained to preset test cases and cannot reproduce real adversarial patterns including autonomous reconnaissance, repeated trial-and-error and composite multi-vector intrusions. Low-interaction honeypots simulate simplified services with limited interaction depth, yielding one-sided attack samples.

2.2 Incomplete Collection of Intermediate Attack Behaviors

Static testbeds only record system states after successful vulnerability exploitation. Probing activities such as port scanning, brute-force trials and unsuccessful exploit attempts are omitted. Traditional honeypots primarily capture network-layer packets, lacking comprehensive logging for application-layer commands, attacker-interaction logic and behavioral sequences. Discarded non-successful intrusion behaviors result in fragmented attack records and loss of forensically valuable details.

2.3 Absence of Automatic Correlation for Multi-Source Datasets

Attack-related data consist of network packets, host logs, operation instructions and timestamps. In conventional systems these datasets are isolated without unified indexing mechanisms. Researchers must manually align timestamps and concatenate discrete events, which incurs high labor overhead and potential subjective bias [6]. Complete and verifiable attack-trace chains cannot be generated automatically, undermining forensic continuity.

2.4 Security Risks in Uncontrolled Simulation Testbeds

Attack-simulation experiments are inherently high-risk. Traditional test environments lack logical-network isolation and real-time behavioral monitoring [7]. Malicious traffic may leak outward and trigger unauthorized access against external systems, violating cybersecurity compliance standards and constraining large-scale multi-scenario data-collection experiments

[8].

3. Overall Technical Design of the Automatic-Capture Traceable Honeypot Platform

Targeting the above-mentioned technical defects, this paper designs an automatic-capture traceable honeypot platform centered on high-interaction deception technology [9]. It integrates multi-source data acquisition, time-series correlation analysis, automatic trace-link generation and isolated-domain management. A three-layer progressive technical architecture is proposed, implementing a closed-loop workflow: attack deception → multi-source raw-data collection → data parsing & correlation → trace-link generation → forensic-evidence export [10].

3.1 Platform Overall Architecture

The platform is divided into three hierarchical layers: bottom simulation-support layer, middle data-processing layer, and top analysis-output layer.

1. Bottom Simulation-Support Layer: The fundamental technical base, containing multi-component high-interaction honeypot modules, network-isolation control units and multi-dimensional behavior-capture components. It simulates diverse Web services, databases and host services, enforces logical isolation between simulation domains and external networks, and transparently collects network-traffic, host-system events and application-layer interaction commands.

2. Middle Data-Processing Layer: Responsible for raw-data cleansing, standardized parsing and multi-source time-series correlation. Network packets, host logs and tool-parameter records are normalized. Timestamps serve as primary keys to align heterogeneous datasets and construct structured time-ordered attack-behavior datasets, supplying inputs for trace-link generation.

3. Top Analysis-Output Layer: Realizes automatic attack-trace reconstruction, visual graph rendering and forensic-evidence export. Based on correlated time-series datasets, complete adversarial procedures covering reconnaissance, probing, intrusion and post-exploitation are parsed. Visual attack-trace diagrams are generated; standardized forensic packages containing raw logs and traffic dumps can be exported for downstream threat analysis.

3.2 Closed-Loop Business Processing Workflow

1. Business-Scenario Deployment: Multiple simulated enterprise-like business topologies are deployed to expose detectable network services for attacker deception.

2. Attack Deception and Raw-Data Acquisition: When adversaries perform port scanning, vulnerability scanning, exploit execution and command operations within simulation domains, underlying capture modules transparently record all events including source IPs, timestamps, commands, tool parameters, traffic payloads and success/failure status of every attempt, regardless of whether intrusions succeed.

3. Multi-Source Data Cleansing and Time-Series Correlation: Noise in heterogeneous raw data is filtered. Network-layer, host-layer and application-layer records are aligned by timestamp keys to build structured attack-behavior datasets.

4. Automatic Attack-Trace-Link Generation: According to chronological sequences and attack-phase logic, events such as reconnaissance, brute-force attempts and exploit operations are chained to generate timestamp-annotated visual attack-trace graphs that restore full intrusion paths.

5. Forensic-Evidence Export: Raw packet captures, host logs, time-series behavior records and structured trace reports are exportable to complete digital-evidence fixation for threat-intelligence research.

3.3 Core Functional-Module Design

3.3.1 Multi-scenario high-interaction business-simulation module

This module supports flexible deployment of Web applications, relational databases and operating-system hosts with configurable vulnerability combinations and service versions. Deep adversarial interactions including repeated trials and multi-round exploitation are permitted to reproduce real-world intrusion patterns and mitigate simulation-fidelity deficiencies.

3.3.2 Multi-dimensional full-volume behavior-capture module

Collection spans network, host and application layers. Network-layer components capture full packets inside simulation domains; host-layer modules record system logs, process invocations and shell-command executions; application-layer modules log HTTP requests, submitted payloads and tool parameters. Both successful exploits and failed probing attempts are persistently stored to guarantee dataset completeness.

3.3.3 Multi-source time-series correlation and trace-reconstruction module

Taking timestamps as core indexes, heterogeneous collected records are fused and associated. Rule-based pattern recognition distinguishes reconnaissance, scanning, brute-force and post-exploitation phases. Discrete behavioral nodes are automatically concatenated into complete chronological attack-trace chains.

3.3.4 Network-domain isolation and behavioral-control module

All honeypot instances are deployed within independent logical simulation domains completely isolated from external networks. Attack traffic is confined internally to prevent outbound leakage. Built-in detection rules identify and intercept high-risk behaviors, logging abnormal operations to guarantee experimental safety and compliance.

4. Experimental Validation and Result Analysis

4.1 Experimental Setup

Two groups of comparative test environments are constructed. The control group adopts a conventional static testbed, while the experimental group deploys the proposed automatic-capture traceable honeypot platform. Identical business configurations and vulnerability sets are applied for both groups. Representative attack sequences including port scanning, password brute-force attacks and Web-vulnerability exploitation with numerous failed probing attempts are reproduced. Two evaluation metrics are defined: attack-behavior-capture completeness and attack-link-restoration integrity.

4.2 Evaluation Metrics Definition

1. Behavior-Capture Completeness: Compute capture rate as the ratio of actually logged behavior entries against total ground-truth attack events throughout intrusion procedures.

2. Attack-Link-Restoration Integrity: Evaluate whether the complete end-to-end intrusion sequence from initial reconnaissance to post-exploitation can be restored; verify retention of probing/failure nodes and automatic generation of chronologically consistent attack-trace chains.

4.3 Experimental Outcomes

For the control-group static testbed: Only behaviors triggered by successful vulnerability exploitation are retained. Intermediate events such as scanning, brute-force trials and failed exploit attempts are entirely lost. The system merely presents final exploitation results. Manual log sorting is mandatory for intrusion-path reconstruction; automatic generation of complete attack-trace chains cannot be achieved.

For the experimental-group honeypot platform: All reconnaissance scans, repeated failure trials, exploit executions and post-exploitation operations are fully captured regardless of success status. Multi-source datasets are time-aligned automatically, yielding complete visual attack-trace graphs annotated with timestamps for every behavioral node. Raw traffic dumps, system logs and standardized trace-analysis reports can be exported to form verifiable forensic datasets.

Experimental results confirm that the proposed solution records full-life-cycle adversarial behaviors including tentative failed attempts and automatically reconstructs attack chains. It resolves critical pain points such as fragmented attack records and labor-intensive manual trace assembly in conventional simulation environments. The isolation-control module effectively blocks outward traffic diffusion and eliminates experimental security hazards.

4.4 Limitations of Current Implementation

This prototype still has several limitations. Simulated scenarios mainly cover conventional service vulnerabilities; simulation coverage for advanced persistent threat (APT) campaigns, complex intra-network lateral movement and customized malicious-code interactions requires enhancement. Current trace-link reconstruction relies primarily on time-series and heuristic rule-based correlation, without incorporating machine-learning-driven intelligent attack-behavior classification. Moreover, data-storage indexing efficiency under high-concurrency massive-attack workloads needs further optimization.

5. Discussion and Future Work

The presented automatic-capture traceable honeypot realizes complete attack-behavior acquisition and automated attack-chain reconstruction by combining high-interaction deception, multi-dimensional logging, time-series correlation and logical-domain isolation. Compared with static testbeds and low-interaction honeypots, this solution preserves

abundant intermediate adversarial trial-and-error behaviors and constructs continuous time-ordered forensic chains. It supplies high-quality experimental datasets for cyber-threat research, attack traceability and adversarial-TTP mining.

Three directions are identified for follow-up improvements:

1. Expand the scenario library by adding APT simulation environments, multi-level intra-network topologies and sophisticated malware-interaction testbeds to improve capture capability for advanced adversarial activities.
2. Introduce machine-learning algorithms to mine massive collected behavioral datasets, implementing automatic attack-behavior classification and threat-actor tool-profiling to enhance intelligent threat-analysis capacity.
3. Optimize massive-dataset storage and indexing architecture to improve data read-write and query performance under long-running large-scale threat-deception experiments.

6. Conclusion

Aiming at technical challenges including incomplete attack-behavior acquisition, lack of automatic attack-chain reconstruction and security risks in traditional simulation environments, this paper proposes a honeypot-based technical scheme for network-attack-behavior capture and traceability. A three-layer technical architecture is established to implement high-fidelity business simulation, multi-source full-volume attack-behavior logging, time-series data correlation, automatic attack-trace reconstruction and isolated simulation-domain control, forming a complete closed-loop covering attack deception, data collection, trace analysis and forensic-evidence output. Comparative experiments validate that the system can preserve full-life-cycle attack behaviors (including unsuccessful probing trials), generate visual attack-trace chains automatically and guarantee forensic-chain continuity while preventing malicious-traffic leakage. This research provides technical support and experimental references for network-intrusion threat perception, attack forensic traceability and adversarial-mechanism analysis.

Acknowledgments

This paper is supported by Yulin Normal University College Students' Innovation and Entrepreneurship Project: WeiyuShield-A High-Interaction Honeypot Platform for Automatic Attacker Information Capture and Traceability. This work was supported by the Office of the Cyberspace Affairs Commission of the CPC Guangxi Zhuang Autonomous Region Committee through the 2025 Guangxi Special Support Program for Innovative Talent Development in Cyberspace Affairs (Project No. 2025WXRC02) and the project entitled "Research on the Tiered and Categorized Education and Training System and Assessment Standards for Cybersecurity Practitioners in Guangxi" (Project No. 2025WXRC03).

References

[1] Cheswick B. An evening with Berferd in which a cracker is lured, endured, and studied//Proceedings of the Winter USENIX Conference. San Francisco: USENIX Association, 1992:20-24.

[2] Franco J, Aris A, Canberk B, et al. A survey of honeypots and honeynets for Internet of Things, industrial Internet of Things, and cyber-physical systems. IEEE Communications Surveys & Tutorials, 2021, 23(4):2351-2383.

[3] Antonakakis M, April T, Bailey M, et al. Understanding the Mirai botnet//26th USENIX Security Symposium. Vancouver: USENIX Association, 2017:1093-1110.

[4] Wang L, Deng J, Tan H, et al. AARF: Autonomous attack response framework for honeypots to enhance interaction based on multi-agent dynamic game. Mathematics, 2024, 12(10):1508.

[5] Hassan W U, Noureddine M A, Datta P, et al. OmegaLog: High-fidelity attack investigation via transparent multi-layer log analysis//Proceedings of the Network and Distributed System Security Symposium. San Diego: Internet Society, 2020.

[6] Javadpour A, Ja'fari F, Taleb T, Shojafar M, Benzaïd C. A comprehensive survey on cyber deception techniques to improve honeypot performance. Computers & Security, 2024, 140:103792.

[7] Matencio Escolar A, Wang Q, Alcaraz Calero J M. Enhancing honeynet-based protection with network slicing for massive Pre-6G IoT Smart Cities deployments. Journal of Network and Computer Applications, 2024, 229:103918.

[8] Pour M, Khoury J, Bou-Harb E. HoneyComb: A darknet-centric proactive deception technique for curating IoT

malware forensic artifacts//2022 IEEE/IFIP Network Operations and Management Symposium. Budapest: IEEE, 2022:1-9.
[9] Heluany J B. PLC honeypots: Enhancing interaction-level assessment. Electronics, 2024, 13(20):4024.
[10] Garfinkel S. Digital forensics research: The next 10 years. Digital Investigation, 2010, 7:S64-S73.