Urban Security and Social Trust: Evaluating the Resilience of Smart City Architectures Through Adversarial Simulation

Nicola Balzano[1], Vita Santa Barletta[2], Danilo Caivano[3,4], Vincenzo Cantatore[1], Luigi Grosso[1], Pietro Loparco[1], Domenico Mastrolonardo[1], Vincenzo Turturro[1] and Leo Ventura[1]

[1]*MNTCRL LABoratory*
[2]*Cyber Academy for Security and Intelligence (CrASte)*
[3]*Software Engineering Research LABoratory (SERLAB)*
[4]*Consorzio Interuniversitario Nazionale per l'Informatica (CINI)*

Abstract

The evolution of urban environments into Smart Cities introduces critical dependencies between physical infrastructure and digital security. While hyper-connectivity optimizes municipal services, it exponentially expands the attack surface, creating direct risks to citizen trust and social stability. This paper analyzes the security posture of smart urban environments using data derived from HackInFest 2025 (Bari), utilizing the event as a high-fidelity simulation environment. By deploying functional mockups of critical infrastructure—specifically ATMs, transport ticketing kiosks, and CCTV systems—we assessed the exploitability of unattended terminals and surveillance networks in a controlled setting. The deployed challenges replicated real-world vulnerability classes: heap exploitation via House of Tangerine and tcache poisoning, use-after-free conditions, File Stream Reliant Oriented Programming (FSROP), insecure deserialization enabling remote code execution, authentication bypass through predictable Pseudo-Random Number Generator (PRNGs) and bit-flipping attacks, and web application vulnerabilities including Relative Path Override leading to Cross-Site Scripting (XSS) and Cross-Site Request Forgery (CSRF) chains. Adopting a dual framework of Urban Security (infrastructure hardening) and Social Security (preservation of public trust), we examine the convergence of IT and OT (Operational Technology) vulnerabilities. Our findings demonstrate that physical interfaces in public spaces represent a neglected vector in smart city architecture, where compromise leads to immediate degradation of service reliability and erosion of citizen confidence in municipal digital services.

Keywords

CTF, Urban Security, Social Security, Cybersecurity, Binary Exploitation, Heap Corruption, Smart City Infrastructure

1. Introduction

The rapid urbanization of the 21st century has catalyzed a fundamental transformation in municipal infrastructure, with cities worldwide adopting networked technologies to optimize service delivery, resource management, and quality of life [1]. This evolution toward "Smart Cities" represents more than mere technological adoption—it constitutes a fundamental restructuring of the relationship between urban infrastructure, digital security, and citizen trust [2, 3]. While hyperconnected systems promise efficiency gains, they simultaneously create critical dependencies where infrastructure compromise directly threatens both operational continuity and social cohesion [4, 5].

1st Workshop on Symbiotic AI for Cyber Social Security: Human–AI Collaboration for Social Resilience and Trustworthy Digital Ecosystems (SAI4CCS 2026), co-located with the 2nd Joint National Conference on Cybersecurity (ITASEC & SERICS 2026), February 09–13, 2026, Cagliari, Italy

n.balzano2@studenti.uniba.it (N. Balzano); vita.barletta@uniba.it (V. S. Barletta); danilo.caivano@uniba.it (D. Caivano); v.cantatore11@studenti.uniba.it (V. Cantatore); l.grosso1@studenti.uniba.it (L. Grosso); p.loparco4@studenti.uniba.it (P. Loparco); d.mastrolonardo2@studenti.uniba.it (D. Mastrolonardo); v.turturro3@studenti.uniba.it (V. Turturro); l.ventura13@studenti.uniba.it (L. Ventura)
https://github.com/nicolabalzano (N. Balzano); https://github.com/PietroLoparco (P. Loparco); https://github.com/DomenicoMastrolonardo (D. Mastrolonardo)
0009-0001-4286-2173 (N. Balzano); 0000-0002-0163-6786 (V. S. Barletta); 0000-0001-5719-7447 (D. Caivano)

© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2026-11-14

CEUR-WS.org/Vol-4329/paper16.pdf

The convergence of Information Technology (IT) and Operational Technology (OT) in urban environments introduces attack surfaces that extend far beyond traditional cybersecurity concerns [6]. Public-facing terminals—ATMs, transport ticketing kiosks, parking payment systems, and civic service portals—represent particularly vulnerable nodes in this architecture [7, 8]. These interfaces, deployed in physically accessible locations with minimal supervision, serve as potential entry points into broader municipal networks while simultaneously representing highly visible symbols of civic functionality. Their compromise carries dual implications: technical degradation of service infrastructure and erosion of public confidence in municipal competence [9, 10].

Current smart city security discourse predominantly emphasizes network architecture, IoT device vulnerabilities, and backend system hardening [11, 12]. However, this focus overlooks a critical attack vector: unattended public terminals that combine physical accessibility, user-facing interfaces, and network connectivity [13]. These systems often run standard operating systems in kiosk mode—a configuration that, despite its ubiquity, has proven consistently vulnerable to escape techniques that grant attackers full system access [14].

This paper examines urban security through the lens of adversarial simulation, specifically analyzing data from HackInFest[1] 2025, a Capture The Flag (CTF) event conducted in Bari, Italy. By deploying functional mockups of critical urban infrastructure—including ATMs, transport ticketing kiosks, and CCTV surveillance systems—within a controlled competitive environment, the event generated empirical data on exploitability patterns, attack vector efficacy, and the time-to-compromise for common terminal configurations [15, 16]. We adopt a dual analytical framework that considers both Urban Security (infrastructure hardening and resilience) and Social Security (preservation of public trust and service reliability) [17].

Therefore, to achieve the research goal, the study is guided by the following research questions:

- Can Capture The Flag events be used as controlled environments to generate empirical data on the resilience of critical urban infrastructure?
- How do technical vulnerabilities in urban infrastructure translate into erosion of public trust (social security), and what are the mechanisms behind this cascade?
- What effort is required to put the infrastructure of a smart city at risk?

The research contributes to smart city security discourse in three ways: (1) it demonstrates the methodological value of CTF events as controlled environments for infrastructure vulnerability assessment, (2) it provides detailed technical analysis of exploitation chains including heap corruption techniques, deserialization attacks, and authentication bypass vulnerabilities, and (3) it examines the relationship between technical vulnerability and public trust degradation, arguing that terminal security represents not merely a technical concern but a social one with direct implications for citizen engagement with digital municipal services.

2. Background and Related Work

The security landscape of smart cities is defined by the complex integration of IT, OT, and public service domains, creating an ”unbounded attack surface” where vulnerabilities in one component can cascade across interdependent networks [10, 18]. While much attention is paid to IoT device hardening and network architecture, the physical interface layer—specifically public-facing terminals—remains a critical yet neglected vector. Research indicates that these terminals, ranging from ticketing kiosks to payment points, often operate on standard PC hardware running legacy operating systems or ”brittle-by-default” configurations that lack robust protection against physical interaction [13, 5]. Despite their critical role, these systems frequently rely on basic kiosk mode restrictions which have proven consistently vulnerable to elementary escape techniques, allowing attackers to access underlying OS functions and potentially pivot laterally into sensitive municipal databases [14, 19].

[1] https://hackinfest.it

The implications of such compromises extend beyond operational disruption to the erosion of "social trust"—a fundamental requirement for the adoption of digital municipal services [9]. Trustworthiness in smart city infrastructure relies on the assurance of service reliability and data privacy; however, the highly visible nature of terminal failures transforms technical glitches into public spectacles of incompetence. Unlike backend breaches, a compromised public kiosk serves as a tangible signal of insecurity, undermining citizen confidence and raising concerns about the safety of collected Personally Identifiable Information (PII) and behavioral data [20, 21]. Consequently, security in this domain is not merely a technical safeguard but a mechanism for preserving the legitimacy of smart city governance [22].

To rigorously assess these vulnerabilities, Capture The Flag (CTF) competitions have emerged as a superior alternative to theoretical auditing. Unlike standard penetration tests, CTF events provide a controlled, adversarial environment that generates empirical data on exploitability, attack patterns, and time-to-compromise metrics [23, 24]. By simulating realistic infrastructure and incentivizing "learning by doing," CTF frameworks allow researchers to observe actual exploitation chains—from initial kiosk escapes to privilege escalation—offering concrete evidence for prioritizing remediation in systems that are otherwise difficult to test safely in production environments [25, 26].

3. Theoretical Framework: Intersecting Urban and Social Security

Understanding smart city vulnerability requires frameworks that transcend purely technical analysis. We adopt a dual theoretical lens examining Urban Security and Social Security as interconnected dimensions of municipal resilience (Figure 1).

Urban Security encompasses the technical hardening and operational resilience of physical and digital infrastructure. This dimension addresses questions of vulnerability discovery, attack surface minimization, defense-in-depth implementation, and incident response capacity. Traditional cybersecurity metrics—time-to-compromise, vulnerability severity scores, exploit complexity—constitute primary evaluation criteria. Urban security analysis focuses on infrastructure protection, service continuity, and the technical capacity to detect, respond to, and recover from attacks [18].

Within urban security, we distinguish between IT and OT vulnerabilities, recognizing that their convergence in smart cities creates unique challenges. IT vulnerabilities—such as those in terminal operating systems, web interfaces, and network protocols—follow familiar cybersecurity patterns with established remediation approaches. OT vulnerabilities affect physical systems: traffic lights, water management, building automation [27]. OT system compromise creates direct physical consequences, potentially endangering public safety beyond mere data exposure or service disruption.

Public terminals occupy a liminal space in this IT/OT convergence. They combine standard IT components (PCs, operating systems, network connectivity) with OT functions (controlling physical dispensers in ATMs, operating ticketing gates, activating surveillance systems). Their compromise can pivot attacks from IT domains into OT infrastructure, enabling adversaries to move from kiosk interfaces into building management systems, surveillance networks, or payment processing infrastructure [13].

Social Security, in our framework, addresses the preservation of public trust, social cohesion, and citizen confidence in municipal institutions. This dimension recognizes that infrastructure security failures carry social costs beyond direct technical impacts [9]. Service disruptions affect citizen quality of life, data breaches violate privacy and dignity, and demonstrated vulnerabilities erode confidence in institutional competence. Social security metrics include citizen satisfaction with services, willingness to adopt digital municipal offerings, perception of government capability, and trust in civic institutions [22].

The intersection of urban and social security occurs when technical vulnerabilities create social consequences. A compromised ATM network represents both an urban security failure (technical breach of financial infrastructure) and social security failure (erosion of trust in banking services, potential financial losses affecting vulnerable populations, public perception of inadequate civic protection). A hacked ticketing kiosk displaying error messages represents minor technical impact but substantial social impact if citizens observe and discuss system unreliability, spreading skepticism through social

networks [20].

This intersectional framework reveals why public-facing terminals warrant particular attention. Unlike backend systems whose compromise may remain invisible to average citizens, terminal failures occur in highly visible public spaces. They create "teachable moments" where citizens directly observe infrastructure vulnerability, forming lasting impressions about municipal capability and care. The symbolic dimension of public infrastructure—its role in demonstrating civic competence—means that visible failures carry disproportionate social impact relative to their direct technical consequences.

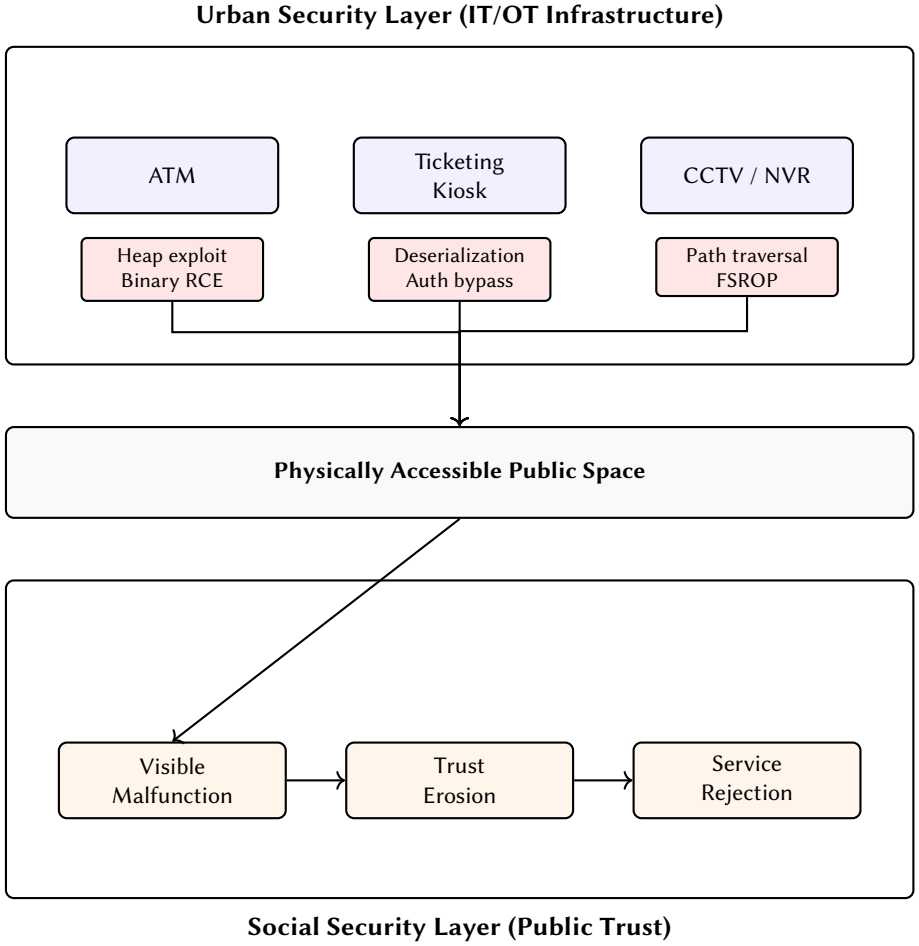


Figure 1: Dual-layer security framework for smart city terminals. Urban security vulnerabilities in physically accessible public terminals (top layer) cascade into social security failures (bottom layer). Technical compromises become visible malfunctions observable by citizens, eroding trust and ultimately leading to service rejection—demonstrating how infrastructure attacks create disproportionate social impacts.

Moreover, the framework highlights how proactive security measures can enhance both urban and social security simultaneously. Adversarial simulation through events like CTF demonstrates municipal commitment to security, transparency about vulnerabilities, and investment in expertise development. When conducted openly and explained to citizens, such initiatives can build trust even while revealing weaknesses—the transparency and proactive stance signal institutional seriousness about protection. Conversely, security-through-obscurity approaches that hide vulnerabilities until exploitation occurs maximize social trust damage when inevitable breaches occur.

4. Case Study: The HackInFest 2025 Initiative

4.1. Event Design and Controlled Environment

HackInFest 2025, conducted November 7–8, 2025, at the Apulia Digital Lab in Bari, Italy, provided a high-fidelity simulation environment for assessing smart city infrastructure vulnerabilities. Organized by the University of Bari's Department of Computer Science in partnership with multiple regional institutions, the event deployed functional mockups of critical urban systems within a controlled competitive setting.

The event adopted a two-track structure addressing different skill levels and threat actor profiles. The Junior Track (November 7) employed a Jeopardy-style CTF format where participants solved discrete challenges of progressive difficulty, focusing on foundational security concepts applicable to smart city contexts [28]. The Senior Track (November 8) deployed realistic scenarios requiring end-to-end exploitation of interconnected systems mirroring actual urban infrastructure dependencies.

Over 300 participants engaged with the challenges, including high school students, university students, technical academy members, and cybersecurity professionals. This demographic diversity proved methodologically valuable, enabling observation of how attackers with varying skill levels, tool familiarity, and domain knowledge approached identical infrastructure vulnerabilities. That advanced high school students successfully competed against professionals in the Senior Track demonstrates that sophisticated attacks on public infrastructure do not require elite expertise—a finding with significant implications for threat modeling.

The event infrastructure simulated a "Mediterranean smart city under attack", deploying interconnected systems representative of actual urban deployments: transport management, edge computing devices, sensor networks, and critically, public-facing terminals. Technical mentors from MNTCRL[2] and SERLAB[3] laboratories provided real-time support, offering a unique research opportunity to observe not only successful exploits but also failed attack attempts, misconceptions about system architecture, and learning trajectories as participants developed exploitation capabilities throughout the event.

The competitive structure incorporated challenge categories spanning the vulnerability landscape relevant to smart cities: Web Exploitation, Cryptography, Reverse Engineering, Binary Exploitation (Pwn), Hardware/IoT, Digital Forensics and Incident Response (DFIR), and Open Source Intelligence (OSINT). This breadth ensured assessment of diverse attack vectors rather than focusing narrowly on single vulnerability classes. Challenges ranged from straightforward configuration errors to complex multi-stage attacks requiring chaining multiple vulnerabilities—representative of real-world exploitation patterns.

4.2. Target Infrastructure: ATMs, Ticketing Kiosks, and CCTV

The HackInFest environment deployed functional replicas of three critical public-facing infrastructure types: Automated Teller Machines (ATMs), transport ticketing kiosks, and Network Video Recorder (NVR) surveillance systems. These systems were selected for their ubiquity in urban environments, their direct citizen interaction, and their representative vulnerability to attack patterns affecting broader terminal classes.

ATM Infrastructure: The simulated ATMs mirrored standard deployments found in urban environments worldwide. Built on standard platforms, these systems employed kiosk mode configurations intended to restrict user interaction to banking functions while preventing access to underlying operating system capabilities. The terminals included standard ATM components: card readers, cash dispensers, receipt printers, and network connectivity for transaction processing. Additionally, a companion Android mobile application and web interface provided alternative access channels, replicating the multi-platform nature of modern banking infrastructure.

[2] https://www.mntcrl.it/
[3] https://serlab.di.uniba.it/

From a security architecture perspective, the ATMs represented the convergence of IT and OT systems characteristic of smart city infrastructure. The IT layer comprised the operating system, banking application software, and network communication protocols. The OT layer included hardware controllers for physical components—the cash dispenser mechanism, card reader electronics, and printer systems. This architecture creates multiple attack surfaces: software vulnerabilities in the OS or banking application, network protocol weaknesses, hardware interface exploits, and critically, the physical accessibility enabling direct terminal manipulation in public spaces [14].

Ticketing Kiosk Infrastructure: Transport ticketing kiosks deployed in the simulation represented systems common in metropolitan transit networks. These terminals employed PC-based architecture running kiosk-mode software intended to restrict functionality to fare purchase, route information, and ticket printing. The kiosk API supported modern features including digital ticket import functionality—mirroring contemporary transit applications that allow users to transfer tickets between devices.

The ticketing kiosks interfaced with backend fare calculation databases, payment processing networks, and potentially integrated with broader transit management systems. Administrative interfaces enabled privileged operations including system diagnostics, configuration changes, and debug functionality accessible through physical NFC card interfaces on the totem hardware.

NVR/CCTV Infrastructure: The Network Video Recorder system represented surveillance infrastructure increasingly deployed in smart city environments for public safety monitoring. The NVR combined embedded firmware for camera management with a web-based user interface for video access and configuration. This dual-interface architecture—low-level firmware interacting with hardware alongside web interfaces for user access—exemplifies the complex attack surface of modern IoT surveillance systems.

Critically, all three terminal types combined high physical accessibility with network connectivity to sensitive backend systems. Deployed in public spaces with minimal supervision, they provide attackers with extended physical access for reconnaissance, tool deployment, and exploitation attempts. This physical accessibility fundamentally differentiates public terminals from secured data center systems or corporate networks with physical access controls [13].

4.3. Adversarial Simulation Rules and Engagement

The HackInFest adversarial simulation operated under carefully structured rules balancing realism with ethical constraints and educational objectives. Participants engaged in "ethical hacking"—authorized security testing within defined boundaries designed to generate useful security insights while preventing actual harm.

Engagement Rules: Participants were explicitly authorized to probe, exploit, and compromise simulated systems within the event infrastructure. All attacks remained confined to event networks, with technical controls preventing any impact on production systems or external networks. This isolation ensured that aggressive testing techniques—brute force attacks, denial-of-service attempts, system crashes—could be employed without real-world consequences, generating data on what motivated attackers might attempt.

Certain techniques were explicitly prohibited to maintain event integrity and focus on skill assessment rather than resource competition. Denial-of-service attacks against competition infrastructure itself were forbidden—while participants could crash individual challenge systems, attacking the central scoring platform or other participants' network access violated rules. This distinction reflects a key insight: the goal was assessing system vulnerability, not participant network attack capabilities.

The simulation dynamics were as follows:

- **Time Pressure:** CTF players operate in a 6-hour window, favoring "noisy" exploits over the "low and slow" stealth characteristic of Advanced Persistent Threats (APTs).
- **Incentives:** CTF players had the opportunity to win a substantial amount of money or other benefits such as training vouchers.
- **Resource Constraints:** Real-world adversaries may possess 0-day exploits not present in a controlled educational setting. Furthermore, while CTF challenges are often designed with an

"intended" solution path, real-world attacks are unconstrained and may leverage methods that deviate significantly from the vulnerability patterns predicted by system designers.

Challenge Structure: Challenges are deployed in tiered difficulty, rewarding both breadth (solving many simpler challenges) and depth (solving complex multi-stage exploits). Points scaled with difficulty, incentivizing participants to attempt challenging problems while ensuring that those with narrower skill sets could still meaningfully participate. This structure generated data across the vulnerability spectrum rather than concentrating solely on either trivial misconfigurations or exotic theoretical weaknesses.

Each solved challenge required participants to retrieve a "flag"—a unique string proving successful exploitation. This artifact-based proof-of-compromise provides verifiable evidence that participants achieved intended objectives rather than merely claiming success. Challenge flags were located at varying system depths: some in easily accessible configuration files (testing basic reconnaissance), others in database records (requiring SQL injection or authentication bypass), still others requiring full system compromise and privilege escalation.

The time constraint—six hours of active competition per track—provided important methodological standardization. Time-to-compromise serves as a critical vulnerability metric: vulnerabilities exploitable within hours by moderately skilled attackers represent high-priority remediation targets, while those requiring days of analysis present lower immediate risk in opportunistic attack scenarios.

5. Technical Analysis: Identified Vulnerabilities

5.1. Common Weaknesses in Simulated Public Terminals

The HackInFest competition revealed systematic vulnerability patterns across simulated public terminals, patterns that align closely with documented weaknesses in deployed smart city infrastructure [14, 13]. Each target system incorporated multiple vulnerability classes requiring distinct exploitation techniques, demonstrating the multi-layered attack surfaces present in real urban infrastructure.

5.1.1. Network Video Recorder (NVR) System

The NVR challenge demonstrated how surveillance infrastructure harbors critical vulnerabilities across both firmware and web interfaces.

Binary Exploitation Chain: The NVR firmware contained an off-by-one vulnerability in the camera description handler. This primitive enabled information leakage from reallocated heap chunks, revealing heap metadata and libc addresses essential for defeating Address Space Layout Randomization (ASLR). Attackers leveraged House of Tangerine techniques [29]—a modernized heap exploitation methodology that achieves arbitrary memory access without requiring explicit `free()` calls—combined with heap feng shui to create a use-after-free (UAF) condition.

Through careful chunk manipulation exploiting the off-by-one to corrupt top chunk metadata, attackers forced the glibc allocator to free the corrupted top chunk into tcache bins. This enabled tcache poisoning [30]—corrupting the forward pointer in freed chunks to redirect subsequent allocations to arbitrary addresses. The final stage employed File Structure Return Oriented Programming (FSROP) on `stderr`, manipulating the FILE structure's vtable pointer to redirect execution flow to the disabled `download_video` function—achieving code execution without requiring stack address leaks.

Web Interface Vulnerability: The NVR's web UI permitted video segment downloads through an HTTP Live Streaming (HLS) (HTTP Live Streaming) endpoint. A flaw in the `stream_hls_segment` handler allowed path traversal through insufficient input validation, enabling non authenticated users to read temporary files containing HLS segments that should have been access-controlled. This bypassed the authentication system's intent to restrict video access to authorized personnel, potentially exposing sensitive surveillance footage.

5.1.2. Automated Teller Machine (ATM)

The ATM challenge presented a multi-layered attack surface spanning binary exploitation, mobile application security, and web frontend vulnerabilities—reflecting the complex ecosystem of modern banking terminals.

Binary Exploitation Chain: The ATM's core banking software contained a logical vulnerability in the `transfer` function that, when combined with information disclosure primitives, enabled complete system compromise. The `withdraw` function leaked binary base addresses, defeating Position Independent Executable (PIE) protections. The `transfer` function provided heap and libc address leaks through crafted transaction parameters.

With these information leaks, attackers achieved tcache poisoning for arbitrary allocations. The binary employed Partial RELRO (Relocation Read-Only), leaving the Global Offset Table (GOT) writable. By allocating a chunk overlapping the GOT and overwriting the `sleep` entry with a libc one-gadget address—a single instruction sequence that spawns a shell when executed—attackers achieved code execution with banking application privileges.

Mobile Application Vulnerabilities: The companion Android APK employed weak obfuscation that, once reversed through standard tools, exposed all API endpoints and authentication mechanisms. Critical flaws included:

- The money transfer endpoint lacked recipient validation, allowing transfers to arbitrary accounts without verification of account ownership or existence.
- No amount validation permitted negative transfer values—effectively enabling attackers to withdraw funds exceeding their account balance by "transferring" negative amounts, which the backend interpreted as incoming funds.
- The APK hardcoded also the not-yet-published website URL, providing attackers with additional attack surface discovery.

Web Frontend Exploitation: The web interface, served as XHTML files using the same API as the mobile application, implemented Content Security Policy (CSP) with DOMPurify for XSS mitigation. However, a Relative Path Override (RPO) vulnerability allowed attackers to manipulate how the browser resolved relative resource paths.

By crafting URLs that altered the browser's base path interpretation, attackers could load XHTML files in contexts where DOMPurify's relative path resolution failed, preventing the sanitization library from loading. This enabled XSS payload injection that bypassed the intended protections. Attackers chained XSS with Cross-Site Request Forgery (CSRF) to auto-elevate privileges: malicious transactions with injection payloads embedded in transaction description fields were sent to bank administrators, who upon viewing the transaction triggered the XSS and unknowingly elevated the attacker's account privileges.

5.1.3. Transport Ticketing Kiosk

The ticketing kiosk demonstrated vulnerabilities across serialization, authentication, and hardware debug interfaces—attack surfaces often overlooked in kiosk security assessments.

Deserialization Attack: The kiosk API supported ticket import functionality using object serialization—mirroring modern transit applications that allow users to transfer digital tickets between devices. Insufficient input validation during deserialization allowed attackers to craft malicious serialized payloads that, when processed by the server, achieved remote code execution.

The deserialization vulnerability could be exploited with varying severity levels: sophisticated payloads spawned interactive shells with application privileges, while simpler exploits enabled ticket generation bypassing payment validation entirely—undermining the transit system's revenue collection model.

Session Token Vulnerabilities: Authentication tokens contained structural weaknesses exploitable through cryptographic attacks:

- An `is_admin` boolean flag within the session token structure was susceptible to bit-flipping attacks. By systematically modifying individual bits in captured tokens and observing server responses, attackers could identify and flip the specific bit controlling administrative privileges.
- The Pseudo-Random Number Generator (PRNG) seeding administrative operation authorization proved predictable. By analyzing token generation patterns, attackers could compute valid authentication values for privileged operations without possessing legitimate credentials.

Hardware Debug Interface: Upon gaining administrative access through the token vulnerabilities, attackers could interact with the physical kiosk interface and its associated NFC hardware. Reverse engineering the kiosk software revealed undocumented debug operations activated through specially crafted NFC transit cards.

By writing specific command codes to NFC cards and scanning them at the physical totem, attackers could trigger debug mode on the public-facing display. This exposed sensitive operational information including internal network addresses, configuration parameters, and diagnostic data—information valuable for planning further attacks or understanding the broader transit system architecture.

5.2. Possible Operational and Social Impact

While quantitative metrics from the CTF provide technical depth, a qualitative assessment reveals the real-world severity (Table 1). The "Impact" here is measured by the potential for cascading failure in civic order.

Threat Vector	Technical Severity	Social Impact	Real-World Consequence
ATM Compromise	Critical	High	Financial panic, service deserts
NVR/CCTV Breach	High	Moderate	Privacy lawsuits, loss of surveillance
Kiosk Manipulation	Moderate	High	Public perception of system decay

Table 1
Qualitative risk analysis of smart city infrastructure threats.

The identified vulnerabilities enable attack vectors with cascading operational and social impacts extending far beyond individual terminal compromise.

Financial Infrastructure Attacks: ATM compromise enables multiple vectors with direct financial and trust implications. Cash theft via dispenser manipulation becomes possible once privilege escalation enables direct communication with cash dispenser hardware. Card data exfiltration through access to transaction logs and memory enables harvesting of sensitive financial information. Service denial through crashed or compromised ATMs creates service deserts particularly impacting cash-dependent populations [14].

Transportation System Disruption: Ticketing kiosk compromise creates operational chaos and revenue loss. Fare evasion facilitation through forged tickets or payment bypass undermines transit revenue models. Backend system penetration using compromised kiosks as pivot points could potentially enable attacks on traffic management infrastructure or scheduling systems.

Surveillance Network Compromise: NVR vulnerabilities enable video feed manipulation (displaying loop recordings to obscure criminal activity), privacy violations through unauthorized surveillance access, and infrastructure reconnaissance providing intelligence about facility layouts and security protocols for planning subsequent attacks [21].

Cascade Effects and Systemic Risk: Individual terminal compromises become critical when attackers leverage them for broader infrastructure penetration. Lateral movement through inadequately segmented networks enables moving from compromised public terminals into backend systems. Persistent access establishment allows compromised terminals to become long-term footholds in municipal networks, enabling extended exploitation given the typically long discovery-to-remediation timelines in municipal environments.

Social Impact Amplification: The public nature of these systems amplifies social consequences. ATM compromises generating unauthorized transactions undermine financial system trust. Ticketing kiosk failures enabling fare evasion at scale threaten transit revenue models and create perceptions of unfairness among paying customers [31]. Surveillance breaches violate citizen privacy expectations. These failures occur at visible public interfaces where citizens directly observe dysfunction, creating lasting negative impressions that propagate through social networks and media coverage [22, 20].

6. Conclusion and Future Directions

This paper has examined smart city infrastructure security through analysis of adversarial simulation at HackInFest 2025, demonstrating that public-facing terminals represent critical yet neglected attack vectors in urban environments. The findings contribute to smart city security discourse in three key areas: methodological innovation in vulnerability assessment, empirical evidence of systematic exploitability through documented attack techniques, and theoretical integration of urban and social security dimensions.

Principal Findings: The HackInFest case study validates theoretical concerns about terminal security through empirical demonstration. The deployed challenges—replicating real-world vulnerability classes including heap exploitation via House of Tangerine and tcache poisoning, FSROP for code execution, insecure deserialization enabling RCE, authentication bypass through predictable PRNGs and bit-flippable tokens, and web application exploitation through RPO-enabled XSS/CSRF chains—showed that attackers can achieve outcomes ranging from information disclosure to complete system compromise.

The systematic nature of identified weaknesses reveals deeper problems than individual misconfigurations. Memory corruption vulnerabilities appeared across multiple targets despite their different purposes, suggesting embedded systems in public infrastructure commonly suffer from memory safety issues. Authentication failures manifested across all three target types, indicating developers consistently underestimate adversarial capabilities in physical access scenarios. The complexity disparity between attack vectors—from trivial API parameter manipulation to sophisticated heap exploitation—suggests uneven security attention across system interfaces.

Recommendations for Practitioners: These findings argue for defense-in-depth approaches including:

- Memory-safe languages for new terminal development to eliminate entire vulnerability classes
- Comprehensive input validation at all interfaces including serialization endpoints
- Cryptographically secure session tokens with authenticated encryption preventing bit-flip attacks
- Removal or strong authentication of debug functionality, particularly hardware interfaces
- Network segmentation preventing lateral movement from compromised public terminals
- Runtime monitoring detecting anomalous behavior patterns indicating exploitation attempts

Recommendations for Policymakers: The results support mandating security baselines for public terminal deployment and sustained investment in municipal cybersecurity capacity. Events like HackInFest demonstrate the value of public-private-academic partnerships for developing regional security expertise. Regulatory frameworks should require security assessment for smart city infrastructure procurement, with particular attention to public-facing terminals combining physical accessibility with network connectivity.

Future Research Directions: This work opens several research trajectories including longitudinal studies tracking terminal security posture evolution over time, comparative analysis across urban contexts and governance models, and sociotechnical research examining citizen perceptions of terminal security [32]. Furthermore, we intend to expand the scope of vulnerability assessment to AI-integrated environments, specifically investigating Adversarial Machine Learning (AML) attacks to evaluate the resilience of algorithmic decision-making systems embedded in smart city infrastructure. The intersection of urban and social security remains paramount. Technical hardening protects infrastructure; protecting visible public interfaces additionally preserves the citizen trust essential for smart city initiatives to

achieve their potential. Events like HackInFest operationalize this principle—converting adversarial knowledge into civic resilience through controlled discovery, systematic analysis, and collaborative remediation.

Symbiotic AI for Continuous Security Assessment: A critical research direction involves operationalizing CTF-derived adversarial knowledge through Symbiotic AI frameworks that enable continuous human-AI collaboration. We propose a four-stage pipeline: (1) *Knowledge Extraction* from CTF telemetry capturing exploit chains, attack patterns, and time-to-compromise metrics; (2) *AI-Assisted Threat Modeling* where machine learning models trained on CTF data predict vulnerability surfaces and synthesize novel exploit chains for new terminal configurations; (3) *Human-in-the-Loop Validation* where security experts verify AI predictions against operational constraints and adversary motivations; and (4) *Adaptive Defense Deployment* automatically adjusting security controls based on validated threat models. This creates a feedback loop where annual CTF events generate training data, AI models predict high-risk configurations in deployed infrastructure, red teams validate predictions, and production telemetry refines models. Unlike episodic penetration testing, this approach transforms HackInFest-style events into continuous learning mechanisms. Critical implementation requirements include explainable AI for accountability, audit trails documenting decision rationale, and public reporting to build citizen trust through demonstrated security competence. A pilot implementation for Apulia region infrastructure would involve: (Phase 1) retrospective analysis of HackInFest 2025 data generating predictions for Bari municipal terminals; (Phase 2) controlled deployment in 3-5 terminals with human validation of all AI recommendations; (Phase 3) full-scale rollout with automated response for validated threat patterns while maintaining human oversight for novel scenarios.

Acknowledgments

This work was partially supported by the following projects: SERICS - “Security and Rights In the CyberSpace - SERICS” (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU; CYBER-PREDICT: Cyber vulnerability ranking prediction by prescription, Avviso “Reti - Sostegno alla ricerca collaborativa”, - Codice Progetto DUQVKW0; Patto territoriale ”Sistema universitario pugliese” – CUP F61B23000370006.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] V. Zavratnik, D. Podjed, J. Trilar, N. Hlebec, A. Kos, E. Stojmenova Duh, Sustainable and community-centred development of smart cities and villages, Sustainability 12 (2020) 3961.

[2] M. T. Baldassarre, V. S. Barletta, D. Caivano, Smart program management in a smart city, in: 2018 AEIT International Annual Conference, 2018, pp. 1–6. doi:10.23919/AEIT.2018.8577379.

[3] V. S. Barletta, D. Caivano, M. De Vincentiis, A. Pal, M. Scalera, Hybrid quantum architecture for smart city security, Journal of Systems and Software 217 (2024) 112161. URL: https://www.sciencedirect.com/science/article/pii/S0164121224002061. doi:https://doi.org/10.1016/j.jss.2024.112161.

[4] C. Ma, Smart city and cyber-security; technologies used, leading challenges and future recommendations, Energy Reports 7 (2021) 7999–8012. URL: https://www.sciencedirect.com/science/article/pii/S2352484721007265. doi:https://doi.org/10.1016/j.egyr.2021.08.124.

[5] C. Cerrudo, Hacking smart cities, in: RSA conference, 2015, pp. 2–18.

[6] V. S. Barletta, M. Calvano, A. Sciacovelli, Cyber social security in multi-domain operations, in: 2024 IEEE International Workshop on Technologies for Defense and Security (TechDefense), 2024, pp. 41–46. doi:10.1109/TechDefense63521.2024.10863352.

[7] P. Buono, G. Desolda, F. Greco, A. Piccinno, Let warnings interrupt the interaction and explain: designing and evaluating phishing email warnings, in: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, 2023, pp. 1–6.

[8] G. Desolda, F. Di Nocera, L. Ferro, R. Lanzilotti, P. Maggi, A. Marrella, Alerting users about phishing attacks, in: International Conference on Human-Computer Interaction, Springer, 2019, pp. 134–148.

[9] C. SCHMIDT, M. MANLEY, Trust in smart city systems: Characteristics and key considerations, Arlington: Cybersecurity and Infrastructure Security Agency (CISA) (2020).

[10] R. Kitchin, M. Dodge, The (in)security of smart cities: Vulnerabilities, risks, mitigation, and prevention, Journal of Urban Technology 26 (2019) 47–65. URL: https://doi.org/10.1080/10630732.2017.1408002. doi:10.1080/10630732.2017.1408002. arXiv:https://doi.org/10.1080/10630732.2017.1408002.

[11] H. Habibzadeh, B. H. Nussbaum, F. Anjomshoa, B. Kantarci, T. Soyata, A survey on cybersecurity, data privacy, and policy issues in cyber-physical system deployments in smart cities, Sustainable Cities and Society 50 (2019) 101660. URL: https://www.sciencedirect.com/science/article/pii/S2210670718316883. doi:https://doi.org/10.1016/j.scs.2019.101660.

[12] K. Kim, I. M. Alshenaifi, S. Ramachandran, J. Kim, T. Zia, A. Almorjan, Cybersecurity and cyber forensics for smart cities: A comprehensive literature review and survey, Sensors 23 (2023). URL: https://www.mdpi.com/1424-8220/23/7/3681. doi:10.3390/s23073681.

[13] AO Kaspersky Lab, ATM/POS Security Assessment Datasheet, Datasheet / Technical Report, Kaspersky Lab, 2019. URL: https://content.kaspersky-labs.com/se/media/en/business-security/enterprise/atm-and-pos-security-assessment-datasheet.pdf, accessed: 2025-12-19.

[14] M. Fraile, M. Ford, O. Gadyatskaya, R. Kumar, M. Stoelinga, R. Trujillo-Rasua, Using attack-defense trees to analyze threats and countermeasures in an atm: a case study, in: IFIP Working Conference on The Practice of Enterprise Modeling, Springer, 2016, pp. 326–334.

[15] C. Pardo, F. J. Pino, F. García, M. Piattini, M. T. Baldassarre, A process for driving the harmonization of models, 2010, p. 51 – 54. doi:10.1145/1961258.1961271.

[16] C. Pardo, F. J. Pino, F. García, M. Piattini, M. T. Baldassarre, S. Lemus, Homogenization, comparison and integration: A harmonizing strategy for the unification of multi-models in the banking sector, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6759 LNCS (2011) 59 – 72. doi:10.1007/978-3-642-21843-9_7.

[17] V. S. Barletta, D. Caivano, C. Catalano, M. de Gemmis, D. Impedovo, Cyber social security education, in: L. T. De Paolis, P. Arpaia, M. Sacco (Eds.), Extended Reality, Springer Nature Switzerland, Cham, 2024, pp. 240–248.

[18] E. Ismagilova, L. Hughes, N. P. Rana, Y. K. Dwivedi, Security, privacy and risks within smart cities: Literature review and development of a smart city interaction framework, Information Systems Frontiers 24 (2022) 393–414.

[19] H. Sukanen, J. Taylor, R. Castaño-Rosa, S. Pelsmakers, T. Lehtinen, T. Kaasalainen, Passive mitigation of overheating in finnish apartments under current and future climates, Indoor and built environment 32 (2023) 1372–1392.

[20] European Cyber Security Organisation (ECSO), ECSO Technical Paper on Cybersecurity Scenarios and Digital Twins, Technical Paper, European Cyber Security Organisation (ECSO), 2023. URL: https://ecs-org.eu/ecso-uploads/2023/07/ECSO_WG6_DigitalTwin-2.1.pdf, version 1.0; accessed 2025-12-19.

[21] N. Kalbo, Y. Mirsky, A. Shabtai, Y. Elovici, The security of ip-based video surveillance systems, Sensors 20 (2020) 4806.

[22] S. Goldsmith, M. Leger, Building citizens' trust to accelerate the smart city movement, 2019. URL: https://datasmart.hks.harvard.edu/news/article/building-citizens-trust-accelerate-smart-city-movement, accessed: 2025-12-19.

[23] J. Vykopal, V. Švábenský, E.-C. Chang, Benefits and pitfalls of using capture the flag games in university courses, in: Proceedings of the 51st ACM Technical Symposium on Computer Science Education, SIGCSE '20, Association for Computing Machinery, New York, NY, USA, 2020, p.

752–758. URL: https://doi.org/10.1145/3328778.3366893. doi:10.1145/3328778.3366893.

[24] V. Švábenský, P. Čeleda, J. Vykopal, S. Brišáková, Cybersecurity knowledge and skills taught in capture the flag challenges, Computers Security 102 (2021) 102154. URL: https://www.sciencedirect.com/science/article/pii/S0167404820304272. doi:https://doi.org/10.1016/j.cose.2020.102154.

[25] J. Werther, M. Zhivich, T. Leek, N. Zeldovich, Experiences in cyber security education: The {MIT} lincoln laboratory {Capture-the-Flag} exercise, in: 4th Workshop on Cyber Security Experimentation and Test (CSET 11), 2011.

[26] M. Andreolini, V. G. Colacino, M. Colajanni, M. Marchetti, A framework for the evaluation of trainee performance in cyber range exercises, Mobile Networks and Applications 25 (2020) 236–247.

[27] D. U. Case, Analysis of the cyber attack on the ukrainian power grid, Electricity information sharing and analysis center (E-ISAC) 388 (2016) 3.

[28] R. Beuran, Capture the flag platforms, in: Cybersecurity Education and Training, Springer, 2025, pp. 193–219.

[29] Shellphish, how2heap: A repository for learning various heap exploitation techniques, 2025. URL: https://github.com/shellphish/how2heap, gitHub repository, accessed 2025-12-19.

[30] SecQuest Information Security LTD, TCACHE Heap Exploitation, White Paper, SecQuest Information Security LTD, 2024. URL: https://www.secquest.co.uk/white-papers/tcache-heap-exploitation, accessed: 2025-12-19.

[31] G. Desolda, A. Esposito, R. Lanzilotti, A. Piccinno, M. F. Costabile, From human-centered to symbiotic artificial intelligence: a focus on medical applications, Multimedia Tools and Applications 84 (2025) 32109–32150.

[32] M. T. Baldassarre, D. Caivano, B. Fernández Nieto, A. Ragone, Ethics-driven incentives: Supporting government policies for responsible artificial intelligence innovation, IEEE Intelligent Systems 40 (2025) 55 – 63. doi:10.1109/MIS.2025.3583222.