



Trusted Hardware, Untrusted Operators: An Insider-Enabled Threat Taxonomy for National Computer-Based Testing Infrastructure

Perangkat Keras Terpercaya, Operator Tak Terpercaya: Taksonomi Ancaman Berbasis Orang-Dalam untuk Infrastruktur Ujian Nasional Berbasis Komputer

Mulkan Fadhli ^{1*}

¹Prodi Teknologi Informasi, Universitas Islam Negeri Ar-Raniry, Indonesia

Corresponden E-Mail: mulkan.fadhli@ar-raniry.ac.id

Makalah: Diterima 27 Juli 2026; Diperbaiki 1 august 2026; Disetujui 4 august 2026
Corresponding Author: Mulkan Fadhli

Abstrak

Literatur keamanan pada ujian berbasis komputer (computer-based testing/CBT) berkembang hampir seluruhnya di sekitar remote proctoring, dengan model ancaman yang mengasumsikan perangkat milik peserta yang tidak terpercaya pada lingkungan yang tidak terkendali. Kerangka ini memperlakukan perangkat milik institusi di bawah pengawasan fisik sebagai kondisi aman. Penelitian ini membantah asumsi tersebut dengan bukti dari Ujian Tulis Berbasis Komputer–Seleksi Nasional Berdasarkan Tes (UTBK-SNBT) Indonesia, sebuah CBT berisiko tinggi yang diikuti 871.496 pendaftar pada 2026 (846.518 hadir) di jaringan pusat ujian perguruan tinggi negeri yang terdistribusi. Berbasis insiden terdokumentasi publik dari penyelenggaraan 2025 dan 2026 — pengungkapan penegak hukum, pernyataan institusi, dan konferensi pers panitia — kami merekonstruksi dua serangan terorganisir yang mengompromikan perangkat ujian milik institusi meskipun terdapat pengawas fisik, detektor logam, dan CCTV. Pada kasus pertama, staf TI kampus dengan hak akses administratif sah memasang perangkat lunak akses jarak jauh; sembilan tersangka dijerat Pasal 30 dan 32 UU ITE. Pada kasus kedua, perangkat proxy tersembunyi berupa dua mini PC, router, dan UPS disembunyikan dalam kardus printer sekitar enam bulan sebelum ujian. Dari insiden-insiden ini kami menurunkan taksonomi lima kelas ancaman dan menyusun model ancaman berbasis STRIDE yang dipetakan ke kerangka MITRE ATT&CK. Kami menunjukkan bahwa satu-satunya kendali yang terbukti efektif terhadap serangan terancang adalah deteksi anomali lalu lintas jaringan, sementara kendali perimeter yang ditekankan dalam wacana publik gagal secara struktural. Kami menyimpulkan bahwa CBT berbasis perangkat terpercaya merupakan model ancaman tersendiri yang belum ter-teorisasi, dan kami menyusun agenda riset untuk kajian sistematisnya.

Keyword: Ancaman Orang-Dalam, Ujian Berbasis Komputer, Pemodelan Ancaman, Infrastruktur Informasi Kritis, STRIDE.

Abstract

The security literature on computer-based testing (CBT) has developed almost entirely around remote proctoring, where the threat model assumes untrusted examinee-owned hardware operating in an uncontrolled environment. This framing treats institution-controlled hardware under physical invigilation as a secure baseline. We argue this assumption is unsound, substantiating the argument with evidence from Indonesia's national university entrance examination (UTBK-SNBT), a high-stakes CBT administered to 871,496 registered candidates in 2026 (846,518 present) across a distributed network of state-university testing centres. Drawing on publicly documented incidents from the 2025 and 2026 administrations, we reconstruct two organised attacks in which institution-controlled examination hardware was compromised despite physical proctors, metal detectors, and closed-circuit surveillance. In the first, campus IT staff with legitimate administrative privileges installed remote-access software on examination workstations; nine suspects were charged under Articles 30 and 32 of the Electronic Information and Transactions Law. In the second, a covert proxy appliance comprising two mini PCs, a router, and an uninterruptible power supply was concealed within a printer carton approximately six months before the examination. From these incidents we derive a five-class threat taxonomy and construct a STRIDE-based threat model mapped to MITRE ATT&CK. We show that, in the documented cases, the control most consistently effective against the most sophisticated attack was network-layer anomaly detection, whereas the perimeter controls emphasised in policy discourse failed by design. We conclude that trusted-hardware CBT constitutes a distinct, under-theorised threat model, and we outline a research agenda for its systematic study.

Keyword: Insider Threat, Computer-Based Testing, Threat Modelling, Critical Information Infrastructure, STRIDE

1. Introduction

High-stakes examinations allocate scarce public goods. When admission to a competitive academic programme is determined by a single test score, the examination ceases to be merely an assessment instrument and becomes a gatekeeping mechanism whose integrity carries substantial economic value. Where that value is high enough, adversaries will invest proportionally to capture it.

Indonesia's UTBK-SNBT exemplifies this condition at national scale. In 2026 the examination registered 871,496 candidates (of whom 846,518 were present) competing for 286,864 places across 145 state institutions — an acceptance rate of 29.42 percent, with 256,369 candidates admitted [1]. Competition is not uniformly distributed: officials report that cheating is overwhelmingly concentrated in a single field. The Minister stated that approximately 99 percent of detected cheating involved candidates for medicine, a claim corroborated by the examination committee chair with reference to the prior year's data.

The economics are correspondingly stark. In 2025, one syndicate was promised IDR 200 million contingent on a candidate's admission to a medical faculty, a success-fee structure rather than advance payment. By 2026, reported fees had risen to IDR 300–700 million per candidate. Against these sums, the surrogate test-taker in the 2025 case received IDR 2 million. The distribution of returns is instructive: the marginal actor is cheap, while value accrues to those controlling access and tooling.

1.1. The Prevailing Threat Model and Its Blind Spot

Academic work on CBT security has been shaped decisively by the COVID-19 transition to remote assessment. The canonical technical study in this area, Burgess et al. [2], reverse-engineered four proctoring suites covering 93 percent of United States law schools, demonstrating that every anti-cheating measure could be trivially bypassed. Their diagnosis is explicit and, within its scope, correct: the fundamental limitation of remote proctoring derives from software executing on hardware the examinee fully controls.

Crucially, that diagnosis carries an implicit corollary. Burgess et al. observe that computerised examinations historically benefited from in-person proctors and institution-controlled hardware, positioning that configuration as the secure condition. The entire remediation logic of the field follows from this: restore institutional control over hardware and environment, and the threat recedes.

The problem with that corollary is that it relocates the trust boundary without examining who sits inside it. Institution-controlled hardware is administered by institutional personnel. If those personnel are themselves the adversary, or are recruited by one, then physical control of the hardware is not a mitigation but an amplifier: the attacker gains privileged, persistent, and unaudited access to the very machines the model presumes trustworthy.

This concern is not merely conceptual. The Indonesian record documents two such compromises within a single examination cycle, in an environment that possessed physical proctors, metal detectors, closed-circuit surveillance, and institution-owned workstations — the full complement of controls the remote-proctoring literature treats as the secure baseline. The evidence therefore permits an empirical test of the assumption rather than only a theoretical objection to it.

Figure 1 makes the structure of the problem explicit. The physical examination perimeter, Boundary 1, is the boundary the remote-proctoring literature seeks to restore, and it remained intact throughout both incidents. Nested inside it, however, lies a second boundary that no prior work examines: the administrative trust extended to the personnel who provision, configure, and maintain the examination estate. Every documented attack originated inside Boundary 1, at Boundary 2, from actors the prevailing model presumes trustworthy.

1.2. Contributions

This paper makes four contributions:

1. We articulate trusted-hardware CBT as a distinct threat model, complementary to and inverted from the remote-proctoring model that dominates the literature.
2. We reconstruct, from open sources, two organised attacks against national examination infrastructure, including a hardware implant with an approximately six-month dwell time, to our knowledge the first such case documented in the CBT security literature.
3. We derive a five-class threat taxonomy for national CBT and construct a STRIDE-based threat model mapped to MITRE ATT&CK techniques [3].
4. We identify a security governance gap, the absence of binding technical standards for examination centres, and show empirically that detection capability currently depends on discretionary institutional capacity.

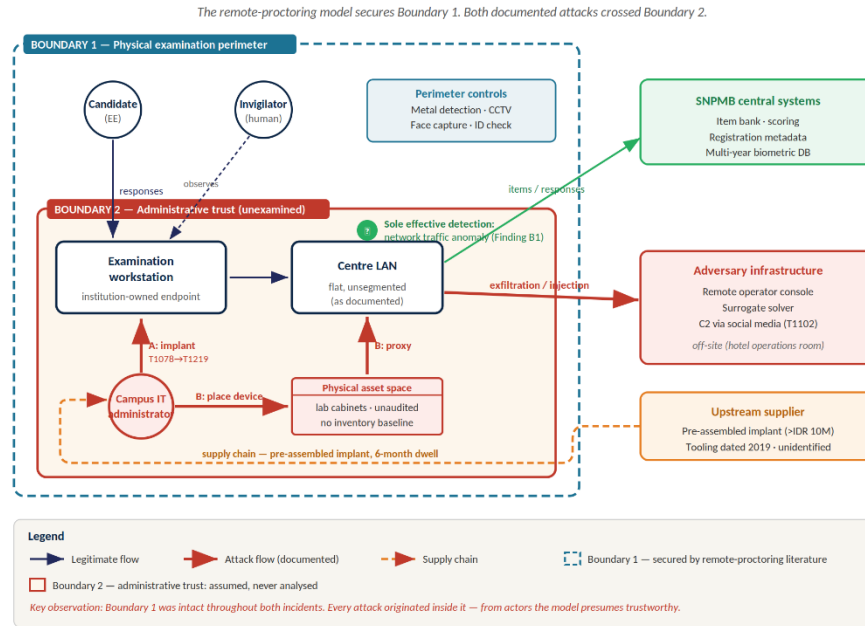


Figure 1. Data flow diagram of the examination environment with trust boundaries. Boundary 1 (physical perimeter) is the locus of remote-proctoring controls and was intact during both incidents. Boundary 2 (administrative trust) is assumed rather than analysed in the existing literature; both documented compromises crossed it.

1.3. Scope and Limitations

This is a case-analysis and position paper based on open-source intelligence. We did not obtain access to system architecture documentation, seized attack artefacts, or non-public incident data. Several details material to a complete technical account, notably the specific remote-access software employed, whether the examination client uses a lockdown browser, and the network topology of testing centres, are not publicly disclosed. This opacity is not merely a constraint on our method; it is itself a finding. A national examination whose security architecture is undocumented in the public record cannot be independently assessed.

Three limitations follow from this open-source basis and should be read as bounding the study's claims. **Data completeness:** the record captures only detected and disclosed incidents; successful, undetected compromises are absent by construction, and the reported four-year operation of one syndicate suggests the undetected population is non-empty. **Technical verification:** we rely on official characterisations of technical events and could not independently examine seized artefacts, system architecture, or network captures, so kill-chain reconstructions are inferential. **Reporting bias:** disclosures are shaped by the priorities of law-enforcement and institutional communications, which may over-represent dramatic modalities and under-report mundane ones. These limitations constrain the strength, though not the direction, of the conclusions drawn.

2. Materials and Method (HEADING 1, 10)

2.1. Remote Proctoring Security

Burgess et al. [2] provide the most rigorous technical treatment to date. Their reverse engineering of Examplify, Exam4, Electronic Blue Book, and ILG Exam360 established that virtual-machine detection universally relies on CPU-vendor block lists trivially defeated by reconfiguration; that process restrictions can be evaded by recompiling open-source software under a permitted process name; and that clipboard protections are ineffective against hardware-based external clipboards. Their facial-recognition analysis found the classifier used by the market-leading suite exhibited a false non-match rate of 11.47 percent, rising above 50 percent at 60-degree facial rotation, with substantial variance across racial groups.

Balash et al. [4] surveyed student privacy and security perceptions of proctoring services, while Langenfeld [5] examined the security and fairness of internet-based proctored assessment and Cohnsey et al. [6] analysed institutional risk in the rapid adoption of remote-teaching software. Slusky [7] catalogued the security-feature claims of twenty proctoring suites. What unites this body of work is its threat model: *the adversary is the examinee, operating on their own hardware, in an environment the institution does not control*. The insider, the administrator of the examination environment, appears in this literature, when at all, only as a defender.

2.2. Cheating Detection in Indonesian CBT Research

Beyond Indonesia, machine-learning approaches to online-exam cheating detection have matured [8]; yet these too assume an individual examinee as the adversary. Indonesian scholarship on examination security has developed along three lines, none of which addresses infrastructure-level threats.

Table 1. Indonesian CBT security research lines and their implicit threat models.

Research line	Representative work	Implicit threat model
Computer-vision behaviour detection	Real-time UTBK cheating-gesture detection using YOLOv8 and optical flow [9]; CCTV-based detection with YOLOv5 [10]	An individual examinee behaving suspiciously in front of a camera
Exam-browser evaluation	Effectiveness analysis of an anti-AI CBT exam browser using mixed methods and perception questionnaires [11]	An examinee attempting to access unauthorised software
CBT system development	Laravel and Electron.js CBT over school intranet with fullscreen lock and tab-switching prevention [12]	An examinee attempting to leave the exam application

Each line presupposes an adversary who is an individual candidate acting alone with consumer-grade means. None models a resourced syndicate, a privileged insider, a hardware implant, or a supply chain. This is a substantial mismatch with the documented empirical reality of the system these methods purport to protect.

2.3. The Gap

We find no published work, Indonesian or international, that (a) constructs a formal threat model for national-scale CBT infrastructure, (b) analyses insider threat in examination administration, (c) performs technical analysis of real-world examination attack artefacts, or (d) proposes a security architecture for distributed examination centres. The most significant examination-security incident in Indonesian history occurred in April–May 2025; as of this writing, more than a year later, it remains unanalysed in the security literature. **This paper begins to close that gap.**

2.4. Research Design

We adopt an explanatory multiple-case design [13], appropriate where the research question is how a phenomenon occurs, the investigator lacks control over events, and the focus is contemporary. Two cases are examined in depth, selected on the criterion of maximum technical variation within the same threat class, supplemented by six further incidents used to establish the breadth of the taxonomy.

2.5. Data Sources

All data derive from open sources, distinguished by three tiers of evidentiary weight: (1) primary, official press briefings and law-enforcement disclosures; (2) secondary, national news-agency reporting; (3) contextual, statutory instruments (Law 11/2008 as amended [14]; Law 27/2022 [15]; Law 12/2012 [16]). This version of the manuscript rests substantially on Tier 2; for publication, the author must download and transcribe the SNPMB press-conference recordings (Tier 1) and cite them directly.

Source selection followed three explicit criteria: proximity to the event (first-hand disclosures by law-enforcement or the host institution were preferred over derivative reporting), institutional credibility (official committees, courts, and named officials over anonymous commentary), and corroborability (only claims traceable to at least two independent sources were retained as findings). Triangulation was applied by cross-checking each material fact across tiers; a claim appearing only in secondary reporting but absent from any primary disclosure was recorded as unverified rather than asserted. Where sources disagreed, the discrepancy was reported explicitly rather than resolved by preference, a procedure whose output is visible in the divergent incident counts reported later under “Incident Statistics and the Problem of Definitional Inconsistency” (Table 6).

2.6. Analytical Procedure

1. Timeline reconstruction: events ordered chronologically per case and cross-checked across at least two independent sources; discrepancies reported rather than resolved by preference.
2. Actor-role mapping: participants mapped to functional roles (initiator, tooling supplier, implanter, operator, surrogate) rather than to identities.
3. Kill-chain decomposition: each case decomposed into preparation, implantation, activation, exfiltration, and injection phases.
4. Taxonomy induction: threat classes induced from the full incident set, then tested for exhaustiveness and mutual exclusivity against each documented incident.

5. Threat modelling: a data flow diagram of the examination environment constructed from public information; STRIDE applied per component; techniques mapped to MITRE ATT&CK Enterprise. The taxonomy was derived through a two-stage coding process. **Open coding** first labelled each documented incident by its primary attack vector, actor type, target component, and the control it defeated, yielding an initial set of descriptive codes. **Axial coding** then grouped these codes into candidate classes by shared vector, and each candidate class was tested against two criteria: exhaustiveness (every documented incident must fall into at least one class) and mutual exclusivity (no incident's primary vector may belong to two classes simultaneously). Classes that overlapped were merged or their boundaries redrawn until both criteria held across the full incident set. The resulting five classes, together with the incident-to-technique mapping that grounds them, are documented in Appendix A so that the derivation can be inspected and replicated.

2.7. Ethical Considerations

Three constraints govern this work. First, no operationally replicable attack detail is published: specific software identities, implant configurations, and bypass procedures are withheld, consistent with responsible-disclosure norms. Second, individuals are referenced only by the initials used in official disclosures; we make no attempt at further identification. Third, our analysis addresses systems and threat classes, not the culpability of named persons, which is a matter for the courts.

2.8. STRIDE Decomposition of the Examination Estate

Applying STRIDE [17] to the four data flow components of Figure 1 yields the threat matrix in Table 2. Each cell records whether the corresponding threat category was documented as exploited in the 2025–2026 record (E), is architecturally plausible but not observed (P), or is not applicable to that component (—). The matrix is the analytical bridge between the incident record and the taxonomy that follows: it locates, per component, exactly where the estate's trust assumptions failed.

Table 2. STRIDE threat matrix for the examination estate. E = documented as exploited (2025–2026); P = plausible, not observed; — = not applicable. Components correspond to Figure 1.

Component	S	T	R	I	D	E
Examination workstation (endpoint)	E	E	E	E	P	E
Centre LAN (flat, unsegmented)	E	E	E	E	P	—
Physical asset space (lab, unaudited)	E	E	E	P	P	E
SNPMB central systems / registration	E	P	P	P	P	P

Three patterns are diagnostically important. First, exploitation clusters on **Tampering, Elevation of Privilege, and Spoofing at the endpoint**, the workstation is the only component with a documented exploit in five of six categories, because it is the node where an administrative insider's legitimate rights (E, elevation) enable code implantation (T, tampering) that then relays content under a genuine candidate's identity (S, spoofing). Second, **Repudiation is exploited across every institution-side component**: no per-action attribution survives when the acting party is a privileged administrator operating through legitimate channels and a fabricated social-media relay, so the incidents leave no reliable audit trail of who did what. Third, **Denial of service is uniformly absent (P, never E)**, the adversary's incentive is covert continuity, not disruption, which is precisely why long-dwell persistence (Case B) rather than sabotage characterises the threat.

The matrix also exposes what the deployed controls miss. The 2026 control set concentrates on Spoofing at the identity layer (facial recognition, card verification), the leftmost column, and the one most visible at the perimeter. Yet the Tampering and Elevation columns, where the endpoint was actually compromised, have no corresponding control: neither endpoint integrity attestation nor privileged-access monitoring appears in the announced posture. This per-component gap is the structural asymmetry later visualised, at the attack-branch level, in Figure 2.

3. Results and Discussion

3.1. Case A: Privilege Abuse by Administrative Insiders

The examination committee's central monitoring team issued an early warning of potential remote-access activity. On 27 April 2025, at approximately 13:00 local time, CCTV footage showed a person apparently installing remote-access software on an examination workstation; invigilators subsequently identified a machine under suspected remote control. Five workstations were ultimately found to carry the implanted application. On 30 April, the surrogate test-taker was apprehended at the examination site. Six suspects were

announced on 7 May 2025; investigation expanded the total to nine by 19 May, with the additional suspects identified as campus IT personnel.

Table 3. Functional role structure, Case A.

Role	Position	Function
Initiator	External organiser	Recruited the surrogate; directed the question-answer relay; received examination content on his own machine
Tooling supplier	External	Supplied the remote-access application, transmitted via an intermediary to the implanters
Implanter	Contract staff, campus IT administrator	Installed the application on examination workstations; also functioned as server administrator
Operator	External	Operated the remote session; received questions; relayed answers
Surrogate	Second-year medical student at the host institution	Solved examination items remotely; compensated IDR 2 million

The decisive property of this attack is that **no software vulnerability was exploited**. The implanter possessed legitimate administrative authority over the target machines. In ATT&CK terms this is Valid Accounts (T1078) enabling Remote Access Software (T1219). No patch would have prevented it; no vulnerability scanner would have detected it. The control that failed was not technical but organisational — the absence of separation of duties, privileged-access monitoring, and endpoint integrity verification for machines entering an examination role.

Seized evidence included a fabricated social-media account used as the relay channel for answers, corresponding to Web Service (T1102), a C2 pattern characteristic of contemporary malware: it exploits a channel that cannot be blocked without disrupting legitimate operations. The institution's UTBK implementation coordinator stated that the syndicate had been operating for four years; if accurate, the 2025 detection represents the failure of an established operation rather than the interception of a novel one.

3.2. Case B: Hardware Implant with Extended Dwell Time

The central committee warned the host institution of potential remote access on 23 April 2025, the first day of the examination window. The institution's ICT coordinator reported detecting anomalous network traffic on that same day, specifically, traffic from candidate workstations transiting to an unexpected proxy. Tracing localised the destination to a faculty ICT laboratory. An attempted apprehension on day one failed: the device was powered down at 16:00. On day two, with the dean's authorisation, the team conducted a search and recovered the apparatus.

Table 4. Properties of the recovered proxy appliance, Case B.

Property	Reported value and significance
Installation date	October 2024 — approximately six months before the April 2025 examination. A dwell time consistent with advanced persistent threat activity, not opportunistic cheating.
Persistence engineering	Inclusion of a UPS indicates design intent for operational continuity across power interruption — the mark of a device meant to survive unattended for extended periods.
Tooling maturity	The application resident on the proxy was dated 2019, implying a toolchain developed and refined over approximately six years.
Capital investment	The assembly was valued at more than IDR 10 million and was received by the placing individual pre-assembled.
Supply chain	That the device arrived ready-built establishes that the placing insider was a delivery vector, not a fabricator. An upstream supplier exists and was not identified.
Insider profile	Contract technician with eight years' tenure in the faculty ICT laboratory — combining legitimate physical access with environmental knowledge sufficient to select an effective concealment site.
Attribution	IP tracing indicated a destination in a different city; the institution stated that further tracing exceeded its authority and required the national communications ministry.

Detection in Case B was achieved through network traffic anomaly analysis. It was *not* achieved by metal detection, physical invigilation, CCTV review, or facial recognition. This is among the most policy-relevant empirical observations available from these incidents, and it runs contrary to the emphasis of both public discourse and the deployed 2026 control set. The institution reported recovering both client and server applications, stated an intention to conduct a controlled simulation to establish the mechanism, and cautioned that the technique was unlikely to be confined to a single institution.

3.3. Threat Taxonomy for National CBT

Inducing from the full incident set across both administrations, we propose five threat classes, constructed to be exhaustive over documented incidents and mutually exclusive at the level of primary attack vector.

Table 5. Five-class threat taxonomy for national computer-based testing, induced from Indonesian incidents 2025–2026.

Class	Vector	Documented instantiation	Control that failed
A — Insider software implant	Abuse of legitimate administrative privilege	Remote-access application installed by campus IT staff on five workstations	Privileged access management; endpoint integrity
B — Insider hardware implant	Physical access plus supply chain	Two mini PCs, router, UPS in a printer carton; six-month dwell	Physical asset audit; network segmentation
C — Physical covert channel	Sub-detection-threshold body-worn devices	Camera-equipped spectacles and microphone; camera beneath headscarf; miniature hearing aid; micro-cameras on braces, belts, buttons	Metal detection — defeated by design, non-metallic devices
D — Identity spoofing	Manipulation of digital identity artefacts	AI-modified registration photographs; four cards one photo with graded likeness; forged ID cards and diplomas; cross-year surrogates	Manual document verification
E — Structured content exfiltration	Content extraction via legitimate candidates	Alleged mobilisation of early-session candidates to map item patterns; screen recording	Rotation across 23+ item sets

Two structural observations follow. First, classes A and B are insider-dependent and are invisible to every control oriented toward the candidate's body or behaviour; they are addressable only at the endpoint and network layers. Second, class C is defined by its engineering against a specific control, the devices are non-metallic precisely because metal detection is the announced countermeasure. This is adversarial adaptation to a published control, and it implies that any static, disclosed perimeter control has a finite effective lifetime.

Figure 2 decomposes the four attack-capable classes into an attack tree following the convention of Schneier [18], with leaf nodes annotated by MITRE ATT&CK technique. Overlaying the deployed 2026 control set onto this tree exposes a structural asymmetry that is the central diagnostic result of this paper: the two branches that actually compromised the system are precisely the two the control set does not address, while defensive investment has concentrated on the candidate-facing branches that perimeter controls can see.

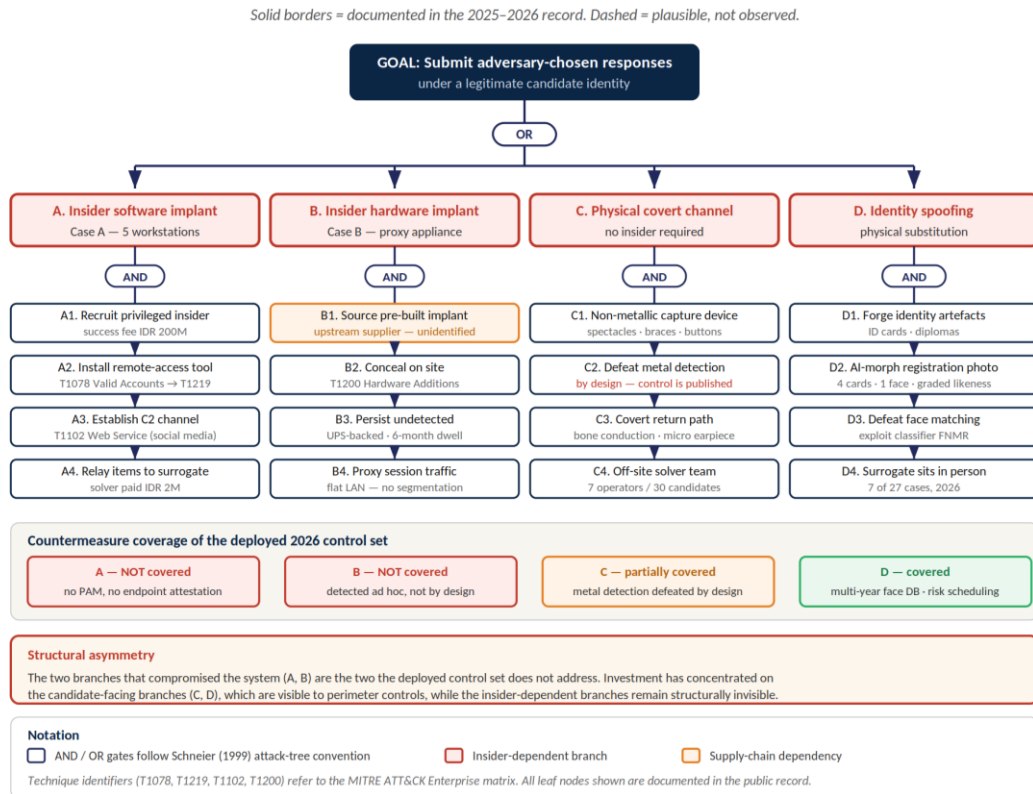


Figure 2. Attack tree for compromise of a national CBT examination session, with countermeasure coverage of the deployed 2026 control set. All leaf nodes are documented in the public record for the 2025–2026 administrations. Technique identifiers refer to the MITRE ATT&CK Enterprise matrix.

3.4. Incident Statistics and the Problem of Definitional Inconsistency

Table 6. Reported incident statistics, UTBK-SNBT 2025–2026.

Year	Candidates	Cheating cases	Surrogates	Notes
2025	860,976	≥50 (as of 29 Apr, sessions 1–12)	10	13 testing centres affected; 14 cases as of 25 Apr
2026	871,496	38 total (27 surrogate + 11 device)	27 (7 present)	1,751 administrative violations; 2,640 pre-exam anomalies

For the 2026 administration the public record contains four incompatible figures: 2,940 anomalous records mapped by the monitoring team (21 April); 2,640 anomalies cited by the Speaker of the House as candidates 'indicated' in cheating (24 April); 1,751 administrative violations; and 38 substantiated cheating cases (25 May). These are not the same category. An anomaly is a detection signal; an administrative violation is a procedural breach; a substantiated case is an adjudicated finding. No public operational definition distinguishes them, and reporting conflates the first with the last, a discrepancy of roughly seventyfold. **The absence of a consistent public incident taxonomy is itself a governance finding.**

3.5. The 2026 Control Set and Its Misalignment

The committee reported a reduction in substantiated cases together with a shift in detection timing from post-examination to pre-examination. The reported control set comprised: AI-based facial recognition against multi-year photograph databases (enabling identification of cross-year repeat surrogates); face capture immediately before room entry; withholding of specific test-centre location at registration; identification of anomalous registration patterns including geographic inconsistency; coded barcodes for card-mismatch detection; metal detection; risk-based scheduling placing medical candidates in the first two days; and location randomisation.

Two observations sit uncomfortably together. First, the 2026 defence rests substantially on facial recognition over a multi-year biometric database — precisely the technology Burgess et al. demonstrate to exhibit high error rates and significant demographic bias. Second, the control that in these cases detected the most sophisticated attack — network anomaly monitoring — does not appear in the announced control set at all. The system has invested against the classes it can see (C and D), while the classes that compromised it (A and B) remain addressed only incidentally.

4. Conclusion and Research Agenda

4.1. Generalizability Beyond UTBK-SNBT

The taxonomy's applicability is bounded by an architectural condition rather than by geography. Classes A and B — insider software and hardware implants — depend on the **institution-controlled-hardware** model, in which the examination runs on machines that an organisation provisions, administers, and physically houses. This condition holds not only for UTBK-SNBT but for any proctored test-centre ecosystem: commercial networks such as Pearson VUE and Prometric, and national examinations administered in institution-operated centres, share the same trust boundary and are therefore susceptible to the same insider-dependent classes. It does **not** hold for at-home remote-proctoring models, where the hardware belongs to the examinee — precisely the configuration the existing literature already addresses. Classes C and D (physical covert channels and identity spoofing) are largely independent of this condition and generalise across both models.

The taxonomy is thus best read as specific to trusted-hardware CBT as a class, of which UTBK-SNBT is one large instance, rather than as either Indonesia-specific or universal to all computer-based testing. Systems such as TOEFL or the GRE, which operate through a mix of test-centre and at-home delivery, would fall under the taxonomy for their centre-based mode and under the remote-proctoring literature for their at-home mode.

4.2. Relation to Established Security Frameworks

The proposed taxonomy is complementary to, rather than a replacement for, established frameworks. Against the **CERT insider-threat typology** (which distinguishes sabotage, theft of intellectual property, and fraud by motive), our classes are organised by **technical vector and target layer** instead of motive, making them directly actionable for control placement; the two views are orthogonal and can be cross-tabulated. Against **STRIDE**, which enumerates threat categories per component (as applied in Section 3.5), our taxonomy supplies the missing empirical layer — which STRIDE categories are actually exercised, and by which concrete attack pattern, in this domain. Against **zero-trust reference models** such as NIST SP 800-207 [19], the taxonomy functions as a threat catalogue that motivates specific zero-trust controls (endpoint attestation for Class A, micro-segmentation for Class B), and Section 6 develops this into a design direction.

We note one limitation of the present validation. The taxonomy has been tested for internal consistency against the documented incident set, but it has not yet been subjected to **external expert review** or applied by independent analysts to a fresh incident corpus. Such validation — for example, a structured expert-elicitation exercise or an inter-rater reliability study of the coding scheme — would strengthen the taxonomy's claim to completeness and is identified as a priority for future work.

We have argued that the CBT security literature rests on an unexamined assumption — that institution-controlled hardware under physical invigilation constitutes a secure baseline — and we have shown, using publicly documented incidents from Indonesia's national entrance examination, that this assumption does not hold. Institution-controlled hardware is administered by people, and where the value at stake reaches hundreds of millions of rupiah per candidate, those people become targets for recruitment — as the two documented cases show.

The inversion is precise. Burgess et al. identify the core problem as software running on hardware the adversary controls. In the UTBK cases, the institution controls the hardware; the adversary controls the administrator. The trust boundary has not been eliminated but relocated, and the relocated boundary was never analysed. This suggests that CBT security requires two threat models, not one.

The state's own prosecution confirms our technical framing: the nine suspects were charged under Article 30(1) illegal access to an electronic system, in conjunction with Article 32(1) and (2) unauthorised alteration and transfer of electronic information, of the Electronic Information and Transactions Law [14], carrying a maximum of nine years' imprisonment. The state treats this as an information-security incident affecting a public electronic system; the security research community has not yet followed.

We propose the following research agenda for the systematic study of trusted-hardware CBT security:

Table 7. Proposed research agenda for trusted-hardware CBT security.

No.	Research direction	Method	Priority
1	Network-layer anomaly detection for insider-enabled remote access	Laboratory replication; attack simulation; feature extraction; unsupervised detection models; precision/recall/F1 evaluation	High — empirically justified
2	Forensic analysis of recovered implant artefacts	Static and dynamic analysis of seized client-server tooling; protocol reconstruction; C2 identification; controlled simulation	Highest value, access-contingent
3	Zero-trust reference architecture for distributed examination centres	NIST SP 800-207-based design [19]; micro-segmentation; pre-session endpoint attestation; cost-feasibility analysis	High
4	Facial recognition efficacy and PDP compliance	Replication of Burgess et al. [2] on Southeast Asian face datasets including headscarf occlusion; Law 27/2022 compliance [16]	High — dual contribution
5	Security capability asymmetry across examination centres	Capability survey; mapping to CIS Controls / ISO 27001; correlation of capability with detection outcomes	Medium — permission-contingent
6	Pre-examination anomaly detection from registration metadata	Feature modelling from geographic inconsistency; risk classification; evaluation against substantiated cases	Medium — data-contingent

The artefacts recovered in Case B, client and server applications from a six-month-dwell implant are, to our knowledge, the only such specimens in existence from a national examination system, and are currently in evidentiary custody. Research access to them would yield the first primary-data forensic account of an attack on national examination infrastructure. That window will not remain open indefinitely.

Appendix A. Incident-to-Technique Mapping

This appendix documents how each documented attack step was mapped to the MITRE ATT&CK Enterprise matrix, providing the reproducible basis for the taxonomy in Section 4.3 and the STRIDE decomposition in Section 3.5. Each row records an observed step, the case it derives from, the assigned technique, and a one-line justification for the assignment.

Table A1. Mapping of documented attack steps to MITRE ATT&CK Enterprise techniques. Case A = Universitas Hasanuddin (2025); Case B = Universitas Jember (2025); Case D = identity-spoofing incidents (2025–2026). Technique identifiers refer to the ATT&CK Enterprise matrix [3].

Observed step	Case	ATT&CK technique	Justification for mapping
Remote-access application installed	A	T1078 Valid Accounts	Access was obtained through legitimate administrative rights, not exploitation; the defining marker of T1078.

Observed step	Case	ATT&CK technique	Justification for mapping
on examination workstations by IT staff with valid credentials			
Installed application used to control the workstation from outside the room	A	T1219 Remote Access Software	A remote-access tool provided interactive control of the endpoint during the session.
Answers relayed through a fabricated social-media account	A	T1102 Web Service	A legitimate web service was used as the command-and-control / exfiltration channel, matching T1102.
Concealed proxy appliance (mini PCs, router, UPS) physically placed in a lab	B	T1200 Hardware Additions	Unauthorised hardware was introduced into the environment to intercept and relay traffic.
Proxy sustained across power interruptions for a six-month dwell	B	T1078 + persistence intent	Persistence was achieved through physical means and privileged placement rather than a software persistence mechanism; recorded as prolonged valid-access abuse.
Candidate workstation traffic transited an unexpected internal proxy	B	T1090 Internal Proxy	Traffic was routed through an adversary-controlled internal node, the definition of T1090.
AI-modified registration photographs and duplicated candidate cards	D	T1036 Masquerading	Identity artefacts were altered to pass verification as a legitimate candidate.

Two mapping decisions warrant comment. First, the six-month dwell of Case B is not itself an ATT&CK technique but a property of the operation; we record it as prolonged abuse of valid access (T1078) enabled by physical placement, since no software persistence mechanism was reported. Second, denial-of-service techniques are absent from the table because none were observed — consistent with an adversary whose objective is covert continuity rather than disruption (Section 3.5).

References

- [1] Panitia SNPMB, “Siaran Pers No. 02/SIPERS/SNPMB/V/2026 tentang Pengumuman Hasil SNBT 2026,” Sistem Nasional Penerimaan Mahasiswa Baru (SNPMB), 2026. [Online]. Available: https://files.snpmb.id/web2026/02_Sipers%20Pengumuman%20SNBT%202026.pdf
- [2] B. Burgess, A. Ginsberg, E. W. Felten, and S. Cohney, “Watching the Watchers: Bias and Vulnerability in Remote Proctoring Software,” arXiv, 2022. doi: 10.48550/arXiv.2205.03009.
- [3] MITRE Corporation, “MITRE ATT&CK® Enterprise Matrix (v17),” 2026. [Online]. Available: <https://attack.mitre.org/matrices/enterprise/>. [Accessed: Jul. 22, 2026].
- [4] D. G. Balash, D. Kim, D. Shaibekova, R. A. Fainchtein, M. Sherr, and A. J. Aviv, “Examining the Examiners’ Privacy and Security Perceptions of Online Proctoring Services,” in Proceedings of the Seventeenth USENIX Symposium on Usable Privacy and Security (SOUPS), USENIX Association, 2021.
- [5] T. Langenfeld, “Internet-Based Proctored Assessment: Security and Fairness Issues,” Educational Measurement: Issues and Practice, vol. 39, no. 3, pp. 24–27, 2020, doi: 10.1111/emip.12359.
- [6] S. Cohney et al., “Virtual Classrooms and Real Harms,” in Proceedings of the Seventeenth USENIX Symposium on Usable Privacy and Security (SOUPS), USENIX Association, 2021.
- [7] L. Slusky, “Cybersecurity of Online Proctoring Systems,” Journal of International Technology and Information Management, vol. 29, no. 1, pp. 56–83, 2020, doi: 10.58729/1941-6679.1421.
- [8] B. Erdem and M. Karabatak, “Cheating Detection in Online Exams Using Deep Learning and Machine Learning,” Applied Sciences, vol. 15, no. 1, p. 400, 2025, doi: 10.3390/app15010400.
- [9] M. N. Ardiansyah, F. Z. Suryahadi, H. E. S. Pratama, and A. P. Sari, “Sistem Deteksi Gerakan Kecurangan UTBK Real-Time dengan YOLOv8 dan Optical Flow,” JISKA (Jurnal Informatika Sunan Kalijaga), vol. 11, no. 1, pp. 83–97, 2026, doi: 10.14421/jiska.5365.
- [10] F. Bimantoro, I. G. P. S. Wijaya, and M. R. Aohana, “Pendeteksian Kecurangan Ujian melalui CCTV Menggunakan Algoritma YOLOv5,” in Seminar Nasional Teknologi & Sains, 2024, pp. 109–117.
- [11] A. Riggo and S. B. Saragih, “Analisis Efektifitas Penggunaan CBT Exam Browser Berbasis Keamanan Digital (Anti AI) untuk Mencegah Kecurangan Ujian UTS/UAS pada Mata Kuliah Manajemen Layanan

- Kesehatan,” *Journal of Innovative and Creativity*, vol. 5, no. 2, pp. 11362–11371, 2025, doi: 10.31004/joecy.v5i2.2790.
- [12] F. N. Iman and I. Farida, “Pengembangan Aplikasi CBT Menggunakan Framework Laravel dan Electron JS,” *JIKA (Jurnal Informatika)*, vol. 9, no. 4, pp. 2713–2722, 2025, doi: 10.31000/jika.v9i4.14795.
- [13] R. K. Yin, *Case Study Research and Applications: Design and Methods*, 6th ed. Thousand Oaks, CA: SAGE Publications, 2018.
- [14] Republic of Indonesia, Law No. 11 of 2008 concerning Electronic Information and Transactions, as amended by Law No. 19 of 2016 and Law No. 1 of 2024. Government of the Republic of Indonesia, 2008.
- [15] Republic of Indonesia, Law No. 27 of 2022 concerning Personal Data Protection. Government of the Republic of Indonesia, 2022.
- [16] Republic of Indonesia, Law No. 12 of 2012 concerning Higher Education. Government of the Republic of Indonesia, 2012.
- [17] A. Shostack, *Threat Modeling: Designing for Security*. Indianapolis, IN: Wiley, 2014.
- [18] B. Schneier, “Attack Trees,” *Dr. Dobbs’s Journal*, vol. 24, no. 12, pp. 21–29, 1999.
- [19] National Institute of Standards and Technology, *Zero Trust Architecture*, NIST Special Publication 800-207. Gaithersburg, MD: U.S. Department of Commerce, 2020, doi: 10.6028/NIST.SP.800-207.