# Measuring and Benchmarking Incident Response Readiness

*Abstract*— **Small-to-medium enterprises (SMEs) remain disproportionately vulnerable to cyber incidents due to constrained resources and underdeveloped operational practices. While many maintain incident response plans (IRPs) to meet regulatory requirements, these plans are often untested and poorly integrated into operational workflows; resulting in delayed containment, unclear escalation, and inconsistent response actions. This disconnect between documentation and execution represents a critical readiness gap that can significantly increase the impact and duration of cyber events. To address this challenge, this paper introduces the Incident Response Readiness Score (IRRS);a scenario-based assessment framework designed to empirically evaluate an organisation's incident response capability under simulated conditions. The IRRS applies a structured scoring rubric calibrated through a Scenario Risk Index, enabling proportional evaluation of performance across diverse incident types. By transforming qualitative incident response actions into a reproducible and risk-weighted metric, the IRRS offers a practical and scalable means of assessing and improving cybersecurity readiness for different type organisations.**

*Keywords— Cybersecurity, incident response evaluation, readiness scoring, Incident Response Planning*

## I. Introduction

Cyber incidents now pose an existential threat for small-to-medium enterprises (SMEs). Despite regulatory frameworks in jurisdictions such as Australia, the United Kingdom, and the European Union mandating breach notification and due diligence obligations regardless of organisational size, SMEs disproportionately struggle to implement robust cybersecurity controls due to inherent resource constraints [1]. With limited access to dedicated security personnel, enterprise-grade tooling, or internal cyber governance functions, SMEs are often forced to prioritise baseline compliance over true operational preparedness. As a result, incident response plans (IRPs) are frequently developed as regulatory artefacts and remain untested; usually archived until a real breach occurs only to find they fail at the first step [2]. This compels urgent improvisation under uncertainty conditions, pressure, and competing business priorities. This gap has been repeatedly cited as a critical vulnerability within the SME sector, particularly given the high prevalence of successful attacks targeting this segment [3]. Major cyber incidents have lead to financial collapse for over 60% of European SMEs, with a majority closing within half a year of the breach [4], Although traditional readiness assessment approaches such as, ISO/ IEC 27035 audits, policy compliance reviews, and tabletop exercises are widespread, current incident response learning processes often fail to integrate people, processes, and technological insights dynamically and instead focusing on limited objectives like service restoration or formal reporting without ensuring broader organisational security learning (5).

Complicating the matter further, leading cybersecurity maturity models such as the NIST Cybersecurity Framework and SIM3 assume availability of substantial security infrastructure, such as Security Operations Centres (SOCs), SIEM platforms, and continuous monitoring regimes resources often beyond the financial or operational capacity of SMEs. While lightweight alternatives such as IASME Cyber Assurance have emerged [6], the broader maturity model landscape remains insufficiently aligned with the live operational realities of SMEs [7]. To address these limitations, this paper introduces the Incident Response Readiness Score (IRRS) a scenario-driven, risk-adjusted framework designed to translate qualitative assessments of IR capability into measurable, repeatable metrics. The IRRS framework operationalises simulations in either live infrastructure or sandboxed digital twin environments to emulate real-world scenarios commonly encountered by SMEs, such as ransomware propagation, cloud storage misconfigurations, insider threats, and business email compromise. This focus on scenario-based validation reflects growing recognition that realistic, time-pressured testing provides deeper insights than checklist-style audits or document-based reviews. IRRS evaluates an organisation's response performance across five interconnected domains: procedural fidelity, operational execution, infrastructure integration, team coordination, and post-incident learning. Each domain is weighted using a Scenario Risk Index (SRI) calibrated to the business impact and complexity of the threat vector under test. Repeated simulation cycles enable longitudinal tracking and benchmarking across a four-tier maturity model; Ad-hoc, Managed, Coordinated, and Adaptive, mirroring progression pathways aligned with SME capability development. The IRRS framework delivers several critical contributions. First, it empowers SMEs to assess and improve their cyber resilience without necessitating enterprise-scale budgets or technologies. Second, it aligns performance measurement with actual organisational behaviours and technical execution, helping SMEs prioritise targeted remediation efforts based on simulation outcomes rather than perceived compliance readiness. Third, early validation studies indicate strong correlation between higher IRRS scores and organisational characteristics such as training frequency, response plan granularity, and team coordination. Finally, by embedding a structured post-incident process, IRRS promotes a culture of continuous improvement through iterative refinement of detection, containment, and recovery capabilities. In sum, IRRS offers SMEs a context-sensitive, defensible, and empirically grounded mechanism to bridge the readiness gap—enabling more informed decision-making and enhancing resilience against increasingly complex cyber threats.

## II. Literature review

The field of incident response (IR) has undergone significant evolution in recent years, transitioning from ad hoc, reactive procedures towards structured, lifecycle-based approaches guided by authoritative frameworks. Notably, the National Institute of Standards and Technology (NIST) Special Publication 800-61 Revision 3 [8], ISO/IEC 27035-2:2023 [9], and the SANS Institute's Incident Handler's Handbook [10] provide comprehensive guidance across critical incident management phases, encompassing preparation, detection, containment, eradication, and recovery. However, despite their widespread recognition, small and medium-sized enterprises (SMEs) commonly face significant barriers to fully operationalising these guidelines due to inherent resource constraints, limited cybersecurity expertise, and informal governance structures [11].

To address these limitations, recent scholarship and industry research have turned towards incident response maturity models (IRMMs), designed to systematically evaluate organisational readiness by benchmarking procedural and technical capabilities. Models such as the Security Incident Management Maturity Model (SIM3) [12] and the European Union Agency for Cybersecurity (ENISA) CSIRT Maturity Framework [13] provide valuable frameworks; however, these tools primarily cater to larger enterprises or national-level teams (e.g., CSIRTs) that possess formalised structures and dedicated response resources. Typically reliant upon qualitative assessments through stakeholder interviews or artefact reviews, these existing maturity models lack empirical rigour, especially in identifying operational bottlenecks and real-world performance discrepancies during live incidents. Such limitations become especially pronounced in SMEs, where incident handling is typically decentralised, ad hoc, and highly context-dependent, thus reducing the efficacy and relevance of conventional IRMM methodologies in these environments.

In parallel, the adoption of simulation-based training and tabletop exercises has gained traction as practical methods to assess and enhance IR capabilities. Advanced simulation tools, such as MITRE ATT&CK Evaluations [14], Caldera [15], and various red/blue team exercises, provide realistic adversary emulation, thereby offering valuable opportunities for organisations to observe and refine their IR processes. Nonetheless, existing simulation tools seldom incorporate structured, quantifiable scoring frameworks to holistically measure IR performance. Furthermore, many of these methods implicitly assume the existence of robust cybersecurity infrastructure including mature security operations centres (SOCs), integrated Security Information and Event Management (SIEM) or Security Orchestration, Automation and Response (SOAR) platforms, and dedicated cybersecurity teams; conditions that rarely exist in SME contexts. Thus, while beneficial in theory, current simulation practices are inadequately tailored to SMEs, limiting their practical utility and leaving a critical gap in accurately measuring and improving IR preparedness within smaller-scale organisations. Moreover, the adoption of risk-adjusted metrics isa practice well established in cybersecurity domains such as vulnerability management (e.g., risk-weighted Common Vulnerability Scoring System [CVSS] methodologies) and compliance assurance (e.g., ISO 27001 audits) remains comparatively unexplored in evaluating incident response effectiveness at an organisational process level [16]. Although maturity models informed by Capability Maturity Model Integration (CMMI) frameworks have occasionally been applied to IR assessments [17], their primary focus continues to be the presence and completeness of documented policies, rather than assessing operational fidelity, execution timeliness, and integration effectiveness during actual or simulated cyber incidents. Recent research has introduced isolated metrics such as time-to-containment (TTC), response success rates, and incident response drift indicators; however, these approaches remain fragmented and lack integration into a unified, comprehensive, and contextually adaptable scoring framework suitable for SMEs. Traditional incident response has often been designed with linear, plan-driven process models that have sequential stages, such as preparation, identification, and containment [16]. Compounding the issue is scant attention has been given in the literature to investigating correlations between IR performance metrics and relevant organisational factors, including team size, tooling integration, and historical incident experience, all of which significantly affect IR outcomes in resource-constrained SME settings. To address the unique cyber resilience challenges faced by SMEs, tailored frameworks and models have been proposed that emphasise prescriptive detection and incident response strategies underpinned by scalable, open-source infrastructure solutions [18]. In parallel, scenario-based training frameworks have been developed to confront the socio-technical barriers that inhibit effective incident response. These meta-level approaches seek to systematically enhance organisational readiness through structured, context-aware exercises that specifically target deficiencies in communication, coordination, and inter-team integration [19].The integration of artificial intelligence (AI) into incident response processes has also been explored, with Oluwawemimo [20] analysing the role of AI in enhancing incident response capabilities within digital domain SMEs. The use of simulation-based training has been highlighted as an effective method for preparing organisations to handle real-world incidents, with studies emphasising the benefits of immersive and scenario-based exercises [21].

This review highlights several persistent gaps in the current body of research that inhibit the effective evaluation of incident response capabilities, particularly within small and medium-sized enterprises. First, there is a notable absence of empirically validated approaches that measure the operational fidelity of incident response execution under realistic conditions. Additionally, current methods lack integrated, risk-adjusted scoring mechanisms that holistically assess both technical controls and procedural efficacy during incident handling. Finally, little attention has been paid to understanding how organisational factors—such as staffing levels, toolchain maturity, and prior incident exposure—shape or influence response outcomes. These omissions collectively limit the development of contextually relevant and practically actionable assessment frameworks for resource-constrained environments. In summary, by providing a rigorous, quantifiable, and contextually tailored approach to incident-response readiness assessment, these studies significantly advance the existing literature and practical capability frameworks. They contribute uniquely by empirically validating IR execution fidelity, linking IR maturity with organisational characteristics, and laying foundational groundwork for future research aimed at optimising incident-response practices specifically within resource-constrained SME contexts.

## III. IRRS Overview and components

This section presents the architectural structure, logic, and scoring methodology of the Incident Response Readiness Score (IRRS). It outlines how IRRS translates simulated incident response behaviour into a reproducible, risk-aware, and analytically supported maturity score for small-to-medium enterprises (SMEs). The framework has been designed for operational realism, evaluative rigour, and actionable outcomes ensuring that incident response capability is measured through performance, not policy alone. The IRRS is a scenario-driven readiness assessment system that enables SMEs to test and evaluate their incident response capability under realistic conditions. Its design is anchored in the execution of simulated incidents such as ransomware attacks, phishing compromises, or insider threats and the observation of organisational behaviour during these events. Rather than simply assessing the presence of documentation (e.g. an IRP) or security tooling, the IRRS evaluates how effectively the organisation applies these resources under stress, scoring across multiple behavioural and technical domains. The end result is a readiness score out of 100 and a mapped maturity tier that reflects the organisation's actual incident response performance.

TABLE I. IRRS Core components

| Component | Function |
|---|---|
| Simulation Engine | Executes realistic, threat-informed incident scenarios (table-top or live emulation) that mirror sector-relevant SME threat profiles and are maintained under version-controlled manifests for reproducibility. |
| Scenario-Specific Evaluation | Assesses participant response to a selected scenario using the Scenario Risk Index (SRI), the Scenario Weighting Guide, and a standardised sub-metric scoring scale (0–5). Integrates context-specific risk weighting with evaluator observations to ensure relevance and proportionality. |
| IRRS Scoring Model | Aggregates scenario scores into a weighted, normalised readiness score (0–100) using predefined sub-metrics across five domains. Designed to enable cross-scenario benchmarking and longitudinal tracking of response capability. |
| Maturity Curve Classification | Tiered classification that maps IRRS scores into discrete readiness bands and surfaces longitudinal trends across repeated simulations. |
| Instrumentation Bus | Passive hooks that record security-tool invocations and human hand-offs during simulations, validating both tooling integration and team-dynamic factors. |
| Feedback & Improvement Actions | Converts simulation findings into targeted remediation guidance across tooling, procedures, and team coordination. Enables iterative readiness uplift and validation of improvement over time. |

The IRRS framework is composed of several interconnected components designed to support realistic simulation, structured evaluation, and actionable insight generation. As outlined in Table I, these include the Simulation Engine, which initiates sector-relevant incident scenarios; Scenario-Specific Evaluation, which contextualises sub-metric scoring using a weighted risk model; and the IRRS Scoring Model, which aggregates performance outcomes. The framework also integrates an Instrumentation Bus to ensure observational fidelity, and a Maturity Curve Classification system that interprets results into actionable tiers. Finally, a Feedback & Improvement mechanism closes the loop, enabling iterative enhancement of readiness practices.

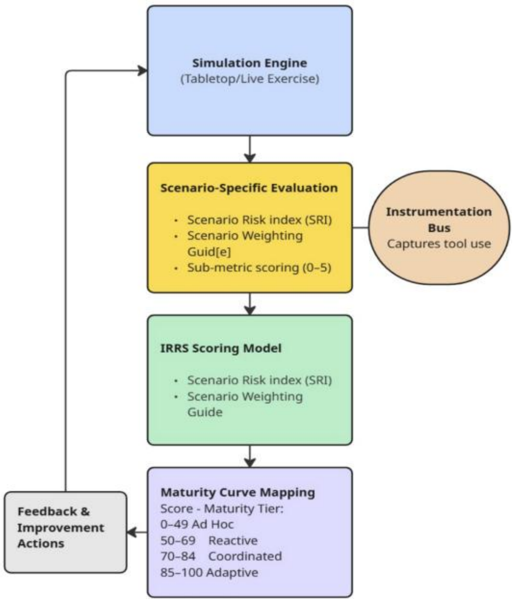

*Figure 1 The IRRS Framework Architecture and Flow*

Figure 1 presents the operational flow of the IRRS framework, illustrating how simulation exercises are translated into measurable readiness insights. The process begins with scenario execution and proceeds through scenario-specific evaluation, risk-adjusted scoring, and maturity tier classification. The Instrumentation Bus enables behavioural data capture throughout, supporting evidence-based assessment. A feedback loop ensures simulation outcomes directly inform iterative improvement efforts. This end-to-end flow reinforces the IRRS model's emphasis on contextual realism, structured evaluation, and continuous organisational uplift.

### A. Simulation Engine

The Simulation Engine is the operational core of the Incident Response Readiness Score (IRRS) framework, generating structured, threat-informed scenarios that replicate the time pressure, ambiguity, and coordination demands of real-world cyber incidents. This controlled environment enables the empirical observation and scoring of incident response behaviours under realistic conditions.

A key advantage of the Simulation Engine is its flexibility. Organisations can select between two modes; tabletop simulation or live emulation depending on their objectives, resource constraints, and maturity level. Tabletop exercises are ideal for assessing procedural alignment, strategic decision-making, and team coordination in a risk-free setting, making them particularly effective for organisations in earlier maturity stages or with limited technical infrastructure. In contrast, live emulations deliver higher-fidelity testing through dynamic telemetry, adversary behaviour, and real-time containment tasks. These are best suited for more mature environments seeking to validate technical controls, detection capabilities, and operational execution under pressure.

By supporting both modes within a unified simulation fabric, the IRRS accommodates a wide range of organisational contexts and ensures that readiness assessments are both scalable and contextually relevant.

TABLE II. SIMULATION MODES AND SCENARIOS

| Mode | Function |
|---|---|
| Tabletop Simulation | Evaluates procedural alignment, strategic decision-making, and team coordination in a risk-free environment. Key artefacts include scenario playbooks, timed inject schedules, facilitator scripts, and observation checklists. |
| Live Emulation | Tests technical detection, containment, and eradication capabilities within isolated virtualised or cloud environments. Artefacts include virtual machines, Infrastructure-as-Code templates, seeded log data, and adversary emulation scripts aligned with MITRE ATT&CK. |

The Simulation Engine supports multiple scenario types, each designed to emulate threat conditions prevalent within small-to-medium enterprises (SMEs). These scenarios enable organisations to assess targeted aspects of their incident response capability under varying levels of risk and complexity. Selection is informed by organisational threat models and tailored to align with sectoral priorities and exposure levels.

### *B. Scenario Specific Scoring & The IRRS Scoring Model*

Following simulation execution, the IRRS framework transitions to structured evaluation through scenario-specific scoring, enabling performance to be quantified relative to the scenario's risk profile and organisational context. To ensure that incident response assessments reflect realistic operational conditions, the framework incorporates a diverse set of scenario archetypes commonly encountered by SMEs. Each scenario type is designed to test distinct aspects of organisational capability, threat exposure, and response behaviour. Scenario selection is informed by threat prevalence, business impact potential, and alignment with resource-constrained operational realities. The IRRS currently supports four representative scenarios: phishing-led ransomware propagation, insider data exfiltration, public-cloud misconfiguration, and credential leakage. Table III summarises these scenario types, highlighting their associated attack vectors and operational implications.

TABLE III. SCENARIO TYPES

| Scenario Type | Details |
|---|---|
| Phishing-led Ransomware Propagation | Simulates a phishing compromise followed by lateral movement and ransomware deployment, resulting in data encryption and extortion. Primary risk vectors include social engineering and endpoint compromise. |
| Insider Data Exfiltration | Models an insider using removable media or personal cloud storage to exfiltrate sensitive data. Risk vectors include insider misuse and unauthorised data access. |
| Public-Cloud Misconfiguration | Emulates exposure of personally identifiable information (PII) due to misconfigured access controls, such as open AWS S3 buckets. Risk vectors include misconfiguration and poor cloud governance. |
| Credential Leakage | Replicates unauthorised access via credentials exposed in public code repositories or compromised CI/CD pipelines. Risk vectors include credential theft and supply chain exposure. |

The Scenario Risk Index (SRI) is a calibrated ordinal scale used within the IRRS framework to weight simulation outcomes according to the operational and strategic significance of the scenario under evaluation. It ensures that readiness assessments reflect not only how well a task was performed, but also how critical that task was within a particular threat context. This approach aligns with risk-adjusted evaluation principles widely adopted in vulnerability scoring systems (e.g., CVSS) and enterprise risk management (ERM) practices. Each scenario is assigned an SRI value ranging from 1 (very low risk) to 5 (very high risk), determined through structured pre-simulation risk modelling.

TABLE IV. IRRS SRI LEVELS

| SRI Level | Risk Description and Representative Scenario |
|---|---|
| SRI 1 – Very Low Risk | Operational impact is negligible. Effects are isolated, non-sensitive, and quickly reversible, with no strategic consequences. Example: benign adware detected; no sensitive data accessed. |
| SRI 2 – Low Risk | Localised impact with limited propagation. May reveal hygiene issues but poses no critical threat. Example: misconfigured antivirus suppresses alerts; minor malware quarantined without incident. |
| SRI 3 – Moderate Risk | Risk of escalation or lateral movement exists. Impact is significant if mishandled, though not immediately urgent. Example: phishing link captures privileged credentials, but account remains unused. |
| SRI 4 – High Risk | Causes business disruption or data exposure. Requires a timely and coordinated response to mitigate legal or reputational consequences. Example: insider exports sensitive customer data via USB. |
| SRI 5 – Very High Risk | Catastrophic scenario involving widespread compromise, system unavailability, or organisational viability risk. Example: ransomware spreads across core servers with exfiltration and extortion. |

This five-level scale, as presented in Table IV, mirrors severity models commonly adopted in operational risk, threat modelling frameworks such as STRIDE, and incident management maturity models. It is deliberately designed as an ordinal scale rather than a ratio-based one for several reasons. First, it emphasises the relative business impact of an incident rather than requiring precise quantification of absolute loss, which is often impractical for SMEs due to limited data and resources. Second, it reflects the triage-based categorisation practices prevalent in many SME environments, where structured classification is more feasible than exhaustive impact measurement. Third, it supports interoperability with detection pipelines, structured playbooks, and prioritisation models used in incident response planning and security operations centre (SOC) workflows.

Within the IRRS framework, each SRI value functions as a weight multiplier, amplifying or moderating the influence of simulation performance depending on scenario severity. Consequently, failures in high-risk simulations (SRI 4–5) are scored more stringently than those in lower-risk contexts (SRI 1–2). This proportional scoring approach not only ensures fairness but also incentivises improved readiness where the potential for harm is greatest. Furthermore, all IRRS simulations are evaluated against fifteen immutable sub-metrics grouped into five readiness domains (Table 2).Fixing the metric catalogue ensures longitudinal comparability across time, teams, and organisations. These sub-metrics form the backbone of IRRS evaluation, ensuring consistency and comparability across scenarios

TABLE V. IRRS SUB-METRICS

| Readiness domain | Sub-metrics |
|---|---|
| Procedural Alignment | • Escalation path followed<br>• IRP referenced during incident<br>• Deviations from IRP formally justified |
| Operational Execution | • Containment-action timing<br>•Task coverage (breadth of technical actions)<br>• Execution accuracy |
| Infrastructure Integration | • Tool-usage effectiveness<br>• Tool alignment to IRP<br>• Inter-tool visibility |
| Co-ordination & Communication | • Role clarity<br>• Decision flow<br>• Communication logging |
| Post-incident Follow-through | • Root-cause analysis<br>• Lessons learned<br>• IRP updated post-simulation |

During a simulation, each of the fifteen sub-metrics shown in Table V is assessed by a designated evaluator based on observed team behaviour. These scores, ranging from 0 (not met) to 5 (fully met), are then combined with scenario-specific weights defined by the SRI to produce a normalised readiness score. This process enables consistent, risk-adjusted evaluation across diverse incident types.

### C. IRRS Scoring Formula

The IRRS scoring process consists of three sequential stages that transform observed behaviours during the simulation into a scenario-weighted, normalised readiness score. The formula is structured to explicitly reflect the integration of evaluator assessments and scenario-derived risk weighting.

$$\text{IRRS} = 100 \times \frac{\sum_{i=1}^{n}(s_i \times w_i)}{5 \times \sum_{i=1}^{n} w_i}$$

TABLE VI. IRRS FORMULA ANALYSIS

| Symbol | Definition | Notes |
|---|---|---|
| $n$ | Total number of universal sub-metrics (fixed at 15) | Ensures every scenario is scored on the same metric set. |
| $s_i$ | Evaluator score for sub-metric $i$ | 0 = not met, 5 = fully met (ordinal scale). |
| $w_i$ | Scenario-specific weight for sub-metric $i$ | 1 – 5, set by the Scenario Risk Index (SRI) table. |
| $5\sum w_i$ | Maximum raw score attainable in this scenario | Multiplies the perfect evaluator mark (5) by the cumulative risk weights; dynamically adjusts to any scenario. |

Computation proceeds in three steps: (1) each sub-metric score $s_i$ is multiplied by its scenario $w_i$ (2) the weighted scores are summed to yield a raw total; (3) the raw total is normalised by the maximum attainable score $5\sum w_i$ and scaled to a [0 – 100] range. This formulation guarantees cross-scenario comparability heavily weighted, high-risk exercises enlarge the denominator in proportion to their severity, so performance is always reported as a true percentage of “risk-adjusted perfection”.

### D. Scenario Weighting Guide & Evaluator Scoring Rubric

To enhance reproducibility and scoring consistency across facilitators, we have developed two supplementary resources available via GitHub [22]:

TABLE VII. SAMPLE EXCERPT FROM SCENARIO WEIGHTING GUIDE

| Scenario Type | Sub-metric | Weight | Justification |
|---|---|---|---|
| Ransomware | Operational Execution – Task coverage | 5 | Broad containment across multiple systems is critical to halting lateral spread. |
| Insider Threat | Operational Execution – Task coverage | 4 | Focus may shift to identity-specific controls and data access channels. |

The above table, which is an sample provides recommended weight values (1–5) for selected IRRS sub-metrics across common scenarios (e.g., ransomware, insider threats). Each weight is justified based on its criticality and role within the incident type.

TABLE VIII. SAMPLE EXCERPT FROM EVALUATOR SCORING RUBIC

| Scenario Type | Score = 0 | Score = 3 | Score = 5 |
|---|---|---|---|
| Task coverage (Ransomware) | Only one host isolated | 5 | Full containment of all compromised assets |
| Root-cause analysis (Credential leak) | No RCA performed | 4 | Multi-layered RCA covering process & tooling |

The above table Offers structured criteria to guide evaluators when scoring organisational behaviour during simulations. Each sub-metric includes behavioural anchors for scoring from 0–5.

To further strengthen reproducibility and evaluation consistency, future iterations of the IRRS framework should include controlled benchmarking exercises involving multiple independent evaluators scoring the same simulation. This would enable measurement of inter-rater reliability and help calibrate scoring expectations across varied facilitator backgrounds. While structured scoring rubrics and scenario weighting guides have been developed to minimise ambiguity, such validation studies would empirically test the framework’s ability to support consistent scoring in realistic, high-pressure settings. Early pilot designs could include cross-sectional scoring of recorded simulations or “double-blind” facilitator exercises within the same SME environment.

### E. IRRS Worked examples

To illustrate how the risk-adjusted formula behaves under different threat contexts, two complete calculations are shown below: Insider-Threat and Ransomware Propagation. The weights come from the framework's scenario matrices; the evaluator scores ($s_i$) are the scores given by the evaluator during the exercise the 0-5 scale. The $s_i w_i$ is the multiplication of the weight by the evaluator score.

TABLE IX. INSIDER THREAT SCENARIO

| Domain | Sub-metric | $s_i$ | $w_i$ | $s_i w_i$ |
|---|---|---|---|---|
| Procedural Alignment | Escalation path followed | 4 | 5 | 20 |
| | IRP referenced during incident | 3 | 5 | 15 |
| | Deviations justified | 4 | 4 | 16 |
| Operational Execution | Containment-action timing | 3 | 4 | 12 |
| | Task coverage | 2 | 4 | 8 |
| | Execution accuracy | 4 | 5 | 20 |
| Infrastructure Integration | Tool-usage effectiveness | 3 | 4 | 12 |
| | Tool alignment to IRP | 2 | 3 | 6 |
| | Inter-tool visibility | 3 | 3 | 9 |
| Co-ordination & Comms | Role clarity | 4 | 5 | 20 |
| | Decision flow | 3 | 5 | 15 |
| | Communication logging | 3 | 4 | 12 |
| Post-incident Follow-through | Root-cause analysis | 3 | 3 | 9 |
| | Lessons learned | 2 | 3 | 6 |
| Procedural Alignment | IRP updated post-simulation | 1 | 3 | 3 |
| Totals | | | $\Sigma w_i$ =60 | $\Sigma s_i w_i$= 183 |

$$\text{IRRS} = 100 \times \frac{183}{5 \times 60} = 61.0$$

TABLE X. RANSOMWARE PROPAGATION SCENARIO

| Domain | Sub-metric | $s_i$ | $w_i$ | $s_i w_i$ |
|---|---|---|---|---|
| Procedural Alignment | Escalation path followed | 4 | 4 | 16 |
| | IRP referenced during incident | 4 | 4 | 16 |
| | Deviations justified | 3 | 3 | 9 |
| Operational Execution | Containment-action timing | 4 | 5 | 20 |
| | Task coverage | 4 | 5 | 20 |
| | Execution accuracy | 4 | 5 | 20 |
| Infrastructure Integration | Tool-usage effectiveness | 3 | 4 | 12 |
| | Tool alignment to IRP | 3 | 3 | 9 |
| | Inter-tool visibility | 3 | 3 | 9 |
| Co-ordination & Comms | Role clarity | 4 | 4 | 16 |
| | Decision flow | 4 | 5 | 20 |
| | Communication logging | 3 | 4 | 12 |
| Post-incident Follow-through | Root-cause analysis | 3 | 3 | 9 |
| | Lessons learned | 2 | 3 | 6 |
| Procedural Alignment | IRP updated post-simulation | 2 | 2 | 4 |
| Totals | | | $\Sigma w_i$ =57 | $\Sigma s_i w_i$= 198 |

$$\text{IRRS} = 100 \times \frac{198}{5 \times 57} \approx 69.5$$

The results illustrate how risk normalisation within the IRRS formula functions effectively. Although the ransomware scenario carries a slightly lower total weight ($\Sigma$ $w_i$ = 57) compared to the insider-threat scenario ($\Sigma$ $w_i$ = 60), the denominator in the IRRS equation scales proportionally, ensuring both outcomes are directly comparable on the same 0–100 scale. The difference in performance between the two simulations is also informative. The team achieved a readiness score of 61 percent in the insider-threat exercise and approximately 69.5 percent in the ransomware simulation. This suggests a comparatively stronger capability in responding to high-intensity, time-critical incidents, particularly where rapid containment and technical execution are essential. Furthermore, the insider-threat score was notably affected by underperformance in task coverage and post-simulation IRP updates; two sub-metrics with weights of 4 and 3, respectively. Despite their moderate weighting, deficiencies in these areas had a measurable impact on the overall readiness score. This underscores that improving lower-scoring, moderately weighted behaviours may yield more significant readiness gains than focusing on marginal improvements in already well-performing sub-metrics.

Together, these comparative calculations validate the IRRS scoring model's fairness and diagnostic utility. It accommodates threat-specific weighting while still producing a unified, interpretable metric that supports benchmarking across scenarios and informs targeted organisational improvement.

## IV. IRRS MATURITY CURVE CLASSIFICATION

The Maturity Curve represents the culminating interpretive layer of the Incident Response Readiness Score (IRRS) framework. Its primary function is to translate an organisation's normalised readiness score into a tiered qualitative assessment, enabling executive-level insight and strategic interpretation of operational performance. By integrating quantitative results from simulation scoring with structured maturity descriptors, the curve facilitates informed decision-making across regulatory reporting, board-level governance, and continuous improvement planning. The Maturity Curve serves three critical purposes. First, it enables comparability by providing a standardised benchmarking mechanism across organisations, industry sectors, and simulation cycles through its fixed interpretive bands. Second, it enhances clarity by abstracting complex, domain-specific metrics into accessible language suitable for non-technical stakeholders, including executives, regulators, and senior leadership. Third, it offers directional value by establishing a developmental trajectory that clarifies the behavioural and procedural advancements required to elevate an organisation's incident response maturity.

The logic underpinning the tiering system is grounded in behavioural indicators and sub-metric patterns observed during simulation exercises. These tiers are deliberately constructed as ordinal maturity stages, reflecting progressively higher levels of procedural adherence, team coordination, tooling integration, and adaptive response capacity.

TABLE XI. IRRS MATURITY TIER MAPPING

| IRRS Score Range | Maturity Tier |
|---|---|
| 0–49 | Ad Hoc |
| 50–69 | Reactive |
| 70–84 | Coordinated |
| 85–100 | Adaptive |

The score ranges were selected based on threshold effects observed in simulation scoring patterns. Organisations scoring below 50 typically demonstrate inconsistent role adherence, informal escalation paths, and unstructured decision-making indicating that incident response efforts are largely improvised. In the 50–69 range, procedural elements may be present but are applied inconsistently, often depending on individual effort rather than coordinated team execution. Scores between 70 and 84 reflect maturing capabilities, with most response actions mapped to defined processes and moderate tooling use, though continuous improvement and integration remain limited. Scores of 85 and above reflect highly structured, proactive, and context-driven response capabilities, supported by regular post-incident learning and real-time decision-making.

This tiering structure ensures that readiness classifications are behaviourally meaningful and proportionally aligned to simulation severity and sub-metric weightings. To assign a maturity tier, the IRRS score is first calculated and normalised to a [0–100] scale. The result is mapped to a tier using the fixed bands shown above. For scores within ±2 points of a boundary (e.g. 68–72), evaluators may apply professional discretion based on observed behaviours and qualitative performance notes. Final classifications are recorded alongside domain-specific feedback and recommended uplift actions.

## V. IRRS INSTRUMENTATION BUS

A key design challenge in scenario-based readiness assessment is ensuring that evaluator scoring reflects observable reality rather than post-hoc rationalisation or facilitator bias. The IRRS framework addresses this challenge through the integration of an Instrumentation Bus; a structured, passive observation mechanism required to be built in or factored into the simulation environment to capture relevant human and system interactions without influencing team behaviour. The objective of this component is to ensure that each sub-metric score is grounded in verifiable evidence relating to procedural alignment, tool invocation, escalation timing, and communication hand-offs. The Instrumentation Bus is not a rigid component but a flexible design layer that must be planned in accordance with simulation scope, available infrastructure, and organisational maturity. In low-technology environments, this may be implemented entirely through facilitator forms and structured observation. In more advanced live emulation contexts, it may include telemetry capture from virtual machines, adversary emulation platforms, or SIEM pipelines. What is essential is not the specific instrumentation mechanism but the deliberate incorporation of observational fidelity into simulation design. Without such planning, IRRS scores risk becoming overly dependent on facilitator inference or participant interpretation. The bus ensures that metrics related to timing, coordination, and procedural adherence are based on verifiable events, thereby strengthening the objectivity, repeatability, and defensibility of readiness assessments. In summary, the Instrumentation Bus is critical for aligning simulation-based assessment with IRRS objectives. It enables scoring to reflect not just whether key actions were taken, but how, when, and by whom? These dimensions are fundamental to understanding the true operational readiness of an organisation under incident pressure.

## VI. FEEDBACK AND IMPROVEMENT ACTIONS

The IRRS framework incorporates a structured feedback mechanism to ensure that insights derived from simulations are translated into actionable improvements. Post-simulation reviews synthesise sub-metric outcomes, behavioural observations, and instrumentation data to identify specific areas of strength and weakness. These findings inform targeted remediation—such as refining escalation paths, enhancing tool integration, or clarifying response roles. Iterative testing using updated scenarios allows organisations to measure progress over time and validate the impact of interventions. This feedback loop reinforces IRRS's core objective: enabling continuous uplift of incident response capability through evidence-based learning. Over successive simulation cycles, tracked improvements can elevate an organisation's IRRS maturity tier—enabling strategic alignment between operational uplift and measurable readiness progression.

## VII. IRRS CASE STUDY

This case study presents a deeply contextualised application of the Incident Response Readiness Score (IRRS) framework, informed by over 15 years of experience in enterprise security operations and SME consulting environments. It is designed to reflect common realities observed across sectors where incident response planning exists only as a policy document, and where response execution during incidents often defaults to informal, reactive behaviours. This case study presents a simulated incident response scenario constructed based on the author's professional experience conducting security reviews, incident response facilitation, and breach post-mortems over a period of 15+ years across SME and enterprise environments. While the structure and scoring of the simulation followed the IRRS methodology rigorously, the scenario and organisational profile have been synthesised from recurring patterns observed in real-world practice. All identifying details have been anonymised or fictionalised, and no client, system, or individual is directly referenced. This case study is therefore not the result of formal human research or organisational data collection, but a composite representation used to demonstrate practical application of the IRRS framework. The methodology aligns with professional practice research norms and does not introduce ethical risk or data privacy concerns.

### A. *Case Study Context and Scenario Overview*

Drawing from extensive experience in breach containment, incident response consulting, and post-incident reviews across multiple sectors, a consistent behavioural pattern has emerged among small-to-medium enterprises (SMEs): although many possess formal Incident Response Plans (IRPs), these documents are frequently treated as compliance artefacts rather than as dynamic operational tools. In live scenarios, actual response behaviour is often governed by individual initiative, typically, from internal IT leads—rather than structured team procedures. Escalation processes are often incomplete or undocumented, creating uncertainty under pressure. Communication is generally informal and unstructured, contributing to delayed containment, and key roles are rarely defined with precision. Furthermore, existing security tools are commonly underutilised due to weak integration into response workflows or a lack of procedural clarity. This case study was designed to simulate such conditions within a realistic scenario and assess organisational performance using the Incident Response Readiness Score (IRRS) framework.

The simulation was structured as a live tabletop exercise based on a credential compromise scenario. In the modelled incident, an external attacker obtained access credentials belonging to a warehouse floor manager and used them to access the internal stock control web interface. The attacker proceeded to extract discount pricing records and successfully escalated privileges by leveraging cached session tokens. Although Microsoft Defender for Endpoint eventually flagged the anomalous activity, no automated containment action was triggered. The purpose of the simulation was to assess the organisation's ability to detect unauthorised access, contain the compromised endpoint, escalate the incident to designated response authorities, and document key decisions while initiating post-incident learning. The participating organisation operates in the retail sector, with integrated warehouse and online storefront operations and a total workforce of approximately 90 staff. Its security stack includes Microsoft 365 E5, Microsoft Defender for Endpoint, Microsoft Cloud App Security, and an outsourced SOC-as-a-Service provider. Known limitations included the absence of SIEM correlation rules for lateral movement, an IRP that had not been updated in over 18 months, and no formal documentation outlining role responsibilities during incident response. Historical patterns of response in the organisation included the initial alert being forwarded to the IT manager, followed by informal coordination via Slack and delayed escalation to director-level decision-makers.

TABLE XII. CASE STUDY OVERVIEW

| Attribute | Details |
|---|---|
| Sector | Retail (warehouse and e-commerce operations) |
| Staff Size | ~90 personnel |
| Security Stack | Microsoft 365 E5, Defender for Endpoint, Cloud App Security |
| Security Operations | Outsourced SOC-as-a-Service |
| Known Gaps | No lateral movement correlation rules; outdated IRP; undefined IR roles |
| Typical Response Pattern | IT-led triage, informal Slack communication, delayed escalation to senior leadership |

### B. *Case Study Simulation and assessment*

The scenario was delivered over a 90-minute facilitated tabletop session coordinated by a Cybersecurity Consultancy. Injects were introduced at structured intervals to simulate the unfolding breach, while evaluators monitored team behaviours and decision-making in real time. For example, at T+5 minutes, a Microsoft Defender for Endpoint alert simulated anomalous login activity from a warehouse subnet, prompting initial triage. At T+15 minutes, synthetic log entries were introduced to represent lateral movement and token replay attempts, testing detection fidelity and escalation pathways. The participant group included the IT Manager (technical lead), Business Systems Lead (identity and access), and Operations Director (executive oversight and escalation). IRRS scoring was performed live using the formal IRRS evaluation rubric, with sub-metrics rated against a credential-compromise-weighted risk matrix. The results were recorded and normalised to a [0–100] scale.

TABLE XIII. CASE STUDY RESULTS

| Evaluation Domain | Score | Maximum |
|---|---|---|
| Procedural Alignment | 12.5 | 25 |
| Operational Execution | 16.2 | 25 |
| Infrastructure Integration | 13.8 | 20 |
| Coordination and Communication | 10.1 | 15 |
| Post-Incident Follow-Through | 10.6 | 15 |
| Total Score (Normalised) | 63.2 | 100 |
| Mapped Maturity Tier | Reactive | |

The assessment revealed several procedural and technical deficiencies. The IRP was not consulted during the response, and containment of the compromised endpoint was delayed by approximately 17 minutes. Escalation to senior decision-makers occurred informally and lacked clear authority handoff. Furthermore, there was no structured post-incident documentation or learning process initiated following the exercise. However, the simulation also highlighted important strengths: the team effectively leveraged Defender alerting, maintained responsive internal communication, and engaged in reflective analysis during the debriefing phase. These findings support the utility of the IRRS framework in revealing operational readiness in practical terms. Despite having modern security tools and a nominal IRP in place, the organisation's behaviours under pressure reflected a reactive maturity level. The case illustrates how simulation-based evaluation, paired with structured scoring, can surface meaningful improvement areas that are often overlooked in policy-centric or audit-based reviews. The Instrumentation Bus element was phased into the credential-compromise simulation. Rather than relying on intrusive telemetry or retrospective participant reporting, the bus was implemented using a combination of structured facilitator injects, timestamped observer logs, and simulation timeline checkpoints. Evaluators documented key behavioural indicators, such as the moment Microsoft Defender for Endpoint generated an alert, the time to initial response, and the path of escalation across functional roles. Escalation to the Operations Director, for instance, was noted along with the elapsed time, communication modality, and any references made to the organisation's IRP. The use of passive observation allowed for real-time assessment of whether containment was timely, whether role responsibilities were clearly understood, and whether tooling was effectively integrated into the decision-making process. While no automated instrumentation was deployed, the design and execution of the simulation enabled the evaluators to reliably score performance across IRRS sub-metrics based on recorded team behaviours, tool interactions, and documented decisions.

## VIII. Limitations and bias mitigation

While the IRRS framework offers a structured and practical method for evaluating SME incident response readiness, certain limitations and potential sources of bias should be acknowledged. One notable source of potential bias lies in evaluator subjectivity. Although structured rubrics and scenario weighting guides are provided, scoring outcomes may still be influenced by the evaluator's interpretation, experience, or professional judgement. Differences in how sub-metrics or scenario-critical actions are perceived can result in inconsistent application of scores across evaluators. This highlights the need for future benchmarking studies involving multiple evaluators to assess inter-rater reliability and improve scoring consistency.

The framework's scope is also limited by scenario coverage. IRRS currently focuses on four representative incident types—ransomware propagation, insider data exfiltration, public-cloud misconfiguration, and credential leakage. While these scenarios reflect prevalent SME threats, they do not cover the full breadth of emerging attack vectors such as supply chain compromise, advanced persistent threats, or zero-day vulnerabilities. As such, findings may not fully generalise to all organisational contexts.

Another risk involves the potential for false confidence. Because the IRRS produces a quantified score and mapped maturity tier, organisations may misinterpret strong performance in one simulation as a proxy for comprehensive readiness. This may result in underappreciation of vulnerabilities not captured in the scenario design. It is therefore recommended that organisations conduct regular, varied simulation cycles and periodically update scenarios to maintain realism and mitigate overconfidence.

Finally, the method may underrepresent latent strengths or weaknesses that do not manifest during a given simulation. Simulation outcomes reflect observable behaviours under controlled conditions, which may not reveal all relevant organisational dynamics. Important capabilities—such as leadership initiative under pressure, undocumented escalation paths, or communication breakdowns—could go unnoticed without broader qualitative inquiry. Repeated simulations, triangulated with post-exercise debriefs and observational data, may help surface these less visible factors. These limitations do not undermine the value of the IRRS framework but rather define a roadmap for future research. Specifically, multi-evaluator benchmarking studies, expansion of scenario libraries, and longitudinal tracking of simulation outcomes across diverse SME environments represent logical next steps in advancing IRRS maturity and reproducibility. The GitHub material [22] provides a foundation for future enhancement and validation of the framework, supporting both customisation and replication to mitigate the limitations identified.

## IX. Discussion

The use of passive observation allowed for real-time assessment of whether containment was timely, whether role responsibilities were clearly understood, and whether tooling was effectively integrated into the decision-making process. While no automated instrumentation was deployed, the design and execution of the simulation enabled the evaluators to reliably score performance across IRRS sub-metrics based on recorded team behaviours, tool interactions, and documented decisions.This study has introduced and evaluated the Incident Response Readiness Score (IRRS), a scenario-driven framework designed to empirically quantify the operational readiness of small-to-medium enterprises (SMEs) in responding to cybersecurity incidents. By integrating structured simulations with a risk-weighted scoring model, the IRRS moves beyond traditional compliance-based assessments to capture the behavioural, procedural, and technical dimensions of incident response capability. The case study demonstrates that critical gaps persist in SMEs' ability to operationalise their incident response plans (IRPs). Despite the presence of documented procedures, the simulation revealed patterns of ad hoc decision-making, delayed containment, and informal escalation; findings that corroborate broader industry observations regarding the static and performative nature of many SME IRPs. The simulation also highlighted underutilisation of existing tools and a lack of structured post-incident documentation, which together limited the organisation's maturity classification to the "Reactive" tier.

Importantly, the application of the IRRS revealed that such shortcomings were not merely anecdotal but could be systematically observed, measured, and scored. While the scoring process necessarily involves some degree of evaluator judgement, the inclusion of a structured rubric, scenario-specific weightings, and real-time behavioural instrumentation (via the Instrumentation Bus) ensures that evaluations are grounded in observable evidence rather than subjective interpretation. The Instrumentation Bus, in particular, was essential to validating the timing, sequence, and delegation of response actions, and its role should be considered a critical design element in any future IRRS-aligned simulation planning. The IRRS framework's most significant contribution lies in its integration of operational realism, behavioural observability, and scenario-sensitive scoring into a cohesive assessment model. The design explicitly accounts for the practical constraints faced by SMEs; such as limited staffing, incomplete toolsets, and decentralised governance while still delivering an interpretable, repeatable readiness metric. Compared to static policy audits or unstructured tabletop exercises, IRRS provides a more empirical, diagnostic, and scalable means of evaluating and improving incident response maturity. The IRRS framework's most significant contribution lies in its ability to translate operational behaviour into measurable readiness scores. However, its long-term value will depend on how well it supports reproducible evaluation, scenario extensibility, and cross-context benchmarking—areas we intend to explore through ongoing deployment and structured facilitator validation efforts.

## X. Conclusion

The Incident Response Readiness Score (IRRS) framework presents a novel, structured and scalable approach for evaluating cybersecurity incident response capability in SMEs through risk-aware, simulation-based assessment. The empirical application of the IRRS in this study confirmed its utility in exposing critical operational deficiencies that may not be apparent in conventional policy-driven evaluations. By combining weighted scoring, structured behavioural observation, and normalised maturity mapping, the IRRS transforms abstract preparedness concepts into measurable, reproducible metrics. It supports maturity benchmarking and organisational uplift by aligning readiness assessments with threat-specific risk profiles and internal capabilities. While the current implementation establishes a strong foundational model, further research is encouraged to test broader scenario libraries, validate IRRS scoring across multiple organisational contexts, and examine how resource constraints affect team coordination, escalation clarity, and response timing. The continued refinement of IRRS has the potential to formalise a sector-standard method for readiness benchmarking; one that is accessible to SMEs yet rigorous enough to guide substantive cybersecurity uplift. In doing so, IRRS offers more than a score: it delivers a structured lens through which organisations can observe, understand, and enhance their real-world incident response behaviours, enabling pragmatic advancement from policy to performance. Future work will focus on broader evaluator benchmarking studies, public scenario library expansion, and integration with SME-scale detection tooling to further strengthen framework adoption and reliability.

## XI. References

[1] Verizon, 2025 Data Breach Investigations Report, Verizon Enterprise, May 2023. [Online].

Available: https://www.verizon.com/business/resources/reports/dbir/

[2] E. Goings, R. Plesco, D. Nides, and D. Kilman, “10 Common cyber incident response mistakes,” KPMG, Apr. 2016. Available: https://assets.kpmg.com/content/dam/kpmg/pdf/2016/04/cyber-incident-response.pdf

[3] Australian Cyber Security Centre (ACSC), Small Business Cyber Security Guide, Australian Signals Directorate, 2021. [Online]. Available: https://www.cyber.gov.au/acsc/view-all-content/publications/small-business-cyber-security-guide

[4] R. Boswell, “60% of European SMEs that are cyber-attacked have to close after six months | Startups Magazine,” Startups Magazine, 2022. https://startupsmagazine.co.uk/article-60-european-smes-are-cyber-attacked-have-close-after-six-months (accessed Mar. 13, 2025).

[5] A. Ahmad, S. B. Maynard, and G. Shanks, “A case analysis of information systems and security incident responses,” International Journal of Information Management, vol. 35, no. 6, pp. 717–723, 2015, doi: https://doi.org/10.1016/j.ijinfomgt.2015.08.001.

[6] Andy Dodd, “How IASME Cyber Essentials Can Protect Your SME,” Adas-ltd.com, Nov. 13, 2024. https://adas-ltd.com/blog/how-iasme-cyber-essentials-can-protect-your-sme/ (accessed May 13, 2025).

[7] A. Chidukwani, S. Zander, and P. Koutsakis, “Cybersecurity preparedness of small to medium businesses: A Western Australia study with broader implications,” Computers & Security, vol. 145, p. 104026, 2024, doi: https://doi.org/10.1016/j.cose.2024.104026.

[8] National Institute of Standards and Technology (NIST), "Computer Security Incident Handling Guide," NIST Special Publication 800-61 Revision 3, 2024. [Online]. Available: https://csrc.nist.gov/publications

[9] International Organization for Standardization, "ISO/IEC 27035-2:2023 – Information technology – Information security incident management – Part 2: Guidelines to plan and prepare for incident response," Geneva, Switzerland: ISO, 2023.

[10] P. Kral, "Incident Handler's Handbook," SANS Institute, 2012. [Online]. Available: https://www.sans.org/white-papers/33901/

[11] R. Banham, “Cybersecurity threats proliferating for midsize and smaller businesses,” Journal of Accountancy, vol. 224, no. 1, 2017.

[12] Open CSIRT Foundation, "SIM3 – Security Incident Management Maturity Model," 2023. [Online]. Available: https://opencsirt.org/csirt-maturity/sim3-online-tool/

[13] European Union Agency for Cybersecurity (ENISA), "CSIRT Maturity Framework – Updated and Improved," 2022. [Online]. Available: https://www.enisa.europa.eu/publications/enisa-csirt-maturity-framework

[14] MITRE Engenuity, "MITRE ATT&CK Evaluations," 2025. [Online]. Available: https://attackevals.mitre-engenuity.org

[15] MITRE Corporation, "CALDERA: Automated Adversary Emulation Platform," 2025. [Online]. Available: https://github.com/mitre/caldera

[16] National Institute of Standards and Technology (NIST), "Common Vulnerability Scoring System v3.1: Specification Document," 2023. [Online]. Available: https://www.first.org/cvss/specification-document

[17] A. Ahmad, S. B. Maynard, K. C. Desouza, J. Kotsias, M. T. Whitty, and R. L. Baskerville, “How can organizations develop situation awareness for incident response: A case study of management practice,” Computers & Security, vol. 101, p. 102122, Feb. 2021, doi: https://doi.org/10.1016/j.cose.2020.102122.

[18] Ilca, Lucian F, O. P. Lucian, and T. C. Balan, “Enhancing CyberResilience for Small and MediumSized Organizations with Prescriptive Malware Analysis, Detection and Response,” Sensors, vol. 23, no. 15, 2023, doi: https://doi.org/10.3390/s23156757.

[19] A. O'Neill, A. Ahmad, and S. Maynard, Cybersecurity Incident Response in Organisations: A Metalevel Framework for Scenario based Training. 2021. doi: https://doi.org/10.48550/arXiv.2108.04996.

[20] E. Oluwawemimo, "The Role of Artificial Intelligence in Incident Response for Digital Domain SMEs," Master's thesis, University of Turku, 2024.

[21] P. J. See, C. Ong, N. Poon, K. H. Soh, S. G. Tan, and S. Dhaliwal, “Scenario based simulation training for incident management: for whom and how,” Policing, vol. 18, p. paae132, Jan. 2024, doi: https://doi.org/10.1093/police/paae132.

[22] M. Abid, “IRRS Supplementary Materials: Scenario Weighting Guide and Evaluator Rubric,” GitHub, 2025. [Online]. Available: https://github.com/tenodex/irrs