SET Framework Extension: Predictive Human Layer Risk Assessment (PHLRA)

Addition to the Social Engineering Testing Framework v2 by Kai Aizen

---

1. Introduction: The Assessment-First Paradigm

Traditional social engineering testing follows a flawed methodology: execute attacks first, assess risk after. This is equivalent to conducting penetration testing before understanding an organization's asset inventory and threat model—a practice no competent security program would endorse for technical assessments.

The Predictive Human Layer Risk Assessment (PHLRA) extension to the SET Framework inverts this model. By leveraging advances in AI-driven behavioral analysis, organizations can now model their human attack surface with the same rigor applied to technical infrastructure.

1.1 The Current Problem

Traditional Approach	Fundamental Flaw
Mass phishing simulations	Tests one vector; misses vishing, pretexting, physical
Click-rate metrics	Measures outcomes, not underlying vulnerability
Point-in-time testing	No predictive capability, only reactive
Uniform testing	Ignores role-based risk differentiation
Individual focus	Misses organizational trust dynamics

1.2 The PHLRA Approach

PHLRA establishes a three-layer model that must be completed *before* any simulation-based testing:

```
Layer 1: Individual Target Modeling (ITM)
Layer 2: Organizational Trust Mapping (OTM)
Layer 3: Predictive Attack Path Analysis (PAPA)
        ↓
Targeted Validation Testing (informed by Layers 1-3)
        ↓
Model Refinement (continuous feedback loop)
```

---

2. Layer 1: Individual Target Modeling (ITM)

2.1 Purpose

Create psychological and behavioral profiles for individuals within the organization to predict susceptibility to specific social engineering tactics.

2.2 Input Sources

Source Type	Data Points	Privacy Consideration
Professional presence	LinkedIn, company bio, publications	Public data only
Communication samples	Email style, Slack patterns (anonymized)	Requires consent/policy
Role metadata	Title, department, reporting line, tenure	HR data, internal
Access level	Systems, data classification, approval authority	IAM/GRC data
Historical performance	Prior simulation results (if available)	Internal metrics

2.3 Output: Individual Risk Profile (IRP)

Each individual receives a structured profile containing:

Psychological Dimensions:

- Big Five personality estimation
- Communication style classification
- Primary motivators and pressure points
- Authority response pattern
- Urgency susceptibility index

Vulnerability Mapping:

- SET tactic susceptibility scores (SET-001 through SET-018)
- Predicted success rates per tactic
- Optimal pretext categories
- Timing and context vulnerabilities

Quantified Metrics:

- Behavioral Risk Score (BRS): 1-100
- Predicted Click-Through Rate (pCTR)
- Predicted Challenge Rate (pCR)
- Estimated Time-to-Report (eTTR)

2.4 BRS Calculation Methodology

The Behavioral Risk Score is calculated using weighted inputs:

```
BRS = Σ(Wi × Fi) / Σ(Wi)

Where:
  Fi = Factor score (normalized 0-100)
  Wi = Factor weight

Factors:
  F1: Role exposure (external-facing, financial authority, admin access)
  F2: Public information availability (OSINT footprint)
  F3: Communication pattern indicators
  F4: Historical simulation performance (if available)
  F5: Psychological susceptibility indicators
  F6: Access level / blast radius potential
```

Default weights (adjustable per organization):

- F1 (Role): 0.20
- F2 (OSINT): 0.15
- F3 (Communication): 0.20
- F4 (Historical): 0.15
- F5 (Psychological): 0.20
- F6 (Access): 0.10

---

3. Layer 2: Organizational Trust Mapping (OTM)

3.1 Purpose

Model how trust, information, and influence flow through the organization. Social engineering exploits trust relationships—understanding these relationships is prerequisite to assessing organizational risk.

3.2 Trust Network Components

Formal Structure:

- Reporting hierarchies
- Department boundaries
- Approval chains
- Committee/group memberships

Informal Structure:

- Communication frequency patterns
- Cross-department collaboration

- Mentorship/influence relationships
- Social clusters

Trust Bridges: Individuals who span multiple trust domains and represent high-value targets:

- Executive assistants (access + trust from leadership)
- IT administrators (technical access + org-wide trust)
- Cross-functional project leads
- Long-tenured employees with broad relationships

3.3 Output: Organizational Trust Graph (OTG)

A weighted directed graph where:

- Nodes = Individuals
- Edges = Trust relationships
- Edge weights = Trust strength (communication frequency, reporting relationship, tenure overlap)
- Node attributes = Individual Risk Profile (from Layer 1)

3.4 Key OTG Metrics

Node-Level:

- Trust Centrality: How many people trust this individual?
- Bridge Score: Does this person connect otherwise separate groups?
- Blast Radius: If compromised, how many downstream targets become vulnerable?

Org-Level:

- Trust Concentration Index: Is trust distributed or concentrated in few individuals?
- Segmentation Score: How well do trust boundaries align with data classification?
- Single Point of Failure Count: Individuals whose compromise cascades broadly

---

4. Layer 3: Predictive Attack Path Analysis (PAPA)

4.1 Purpose

Using ITM profiles and the OTG, model likely attack paths before any testing occurs. Identify which individuals, compromised in which sequence, lead to which organizational impacts.

4.2 Attack Path Components

Entry Points: Rank individuals by:

- BRS (susceptibility)

- Public exposure (targetability)
- Predicted success rate for initial contact

Pivot Opportunities: From each potential entry point, model:

- Who trusts this person?
- What access do downstream targets have?
- What pretexts become available post-compromise?

Terminal Objectives: Map paths to organizational crown jewels:

- Financial authorization
- Sensitive data access
- System administration
- Executive impersonation capability

4.3 Output: Attack Path Matrix

Entry Point	Tactic	Success Prob	Pivot Target	Pivot Pretext	Terminal Objective	Path Risk Score
EA to CFO	SET-006-02	68%	CFO	"Per [EA], you approved..."	Wire authorization	Critical
Jr. Dev	SET-002-02	71%	DevOps Lead	Internal ticket reference	Prod credentials	High
Recruiter	SET-011-04	54%	HR Director	"Candidate follow-up"	Employee PII	High

4.4 Path Risk Scoring

```
Path Risk = P(entry) × P(pivot) × Impact(terminal) × (1 - Detection_probability)

Where:
  P(entry) = Predicted success rate for initial compromise
  P(pivot) = Predicted success rate for lateral movement
  Impact(terminal) = Business impact score of terminal objective
  Detection_probability = Likelihood of reporting/intervention at any stage
```

---

5. Validation Testing Protocol

5.1 Test Selection Criteria

After PHLRA modeling, validation testing targets:

1. **High-confidence predictions**: Test paths where model predicts >70% success
2. **Critical paths**: Test paths leading to crown jewels regardless of probability
3. **Model calibration**: Sample of medium/low probability paths to validate model accuracy

5.2 Test Design

Simulations are designed to validate specific model predictions:

Prediction to Validate	Test Design
Individual susceptible to authority pretext	Targeted SET-001-01 simulation
Trust bridge exploitable	Simulated compromise notification from trusted source
Attack path viable	Multi-stage simulation following predicted path

5.3 Model Refinement

Post-testing, results feed back into the model:

- Adjust BRS calculation weights based on prediction accuracy
- Update trust graph based on observed reporting patterns
- Recalibrate tactic success rate predictions

---

6. Organizational Resilience Index (ORI)

6.1 Purpose

Provide a single, trackable metric for organizational human-layer security posture.

6.2 ORI Components

```
ORI = (1 - Vulnerability_Score) × Detection_Score × Response_Score

Where:
  Vulnerability_Score = Weighted average of individual BRS scores (role-weighted)
  Detection_Score = Aggregate predicted challenge/report rate
  Response_Score = Organizational response effectiveness (escalation speed, containment)
```

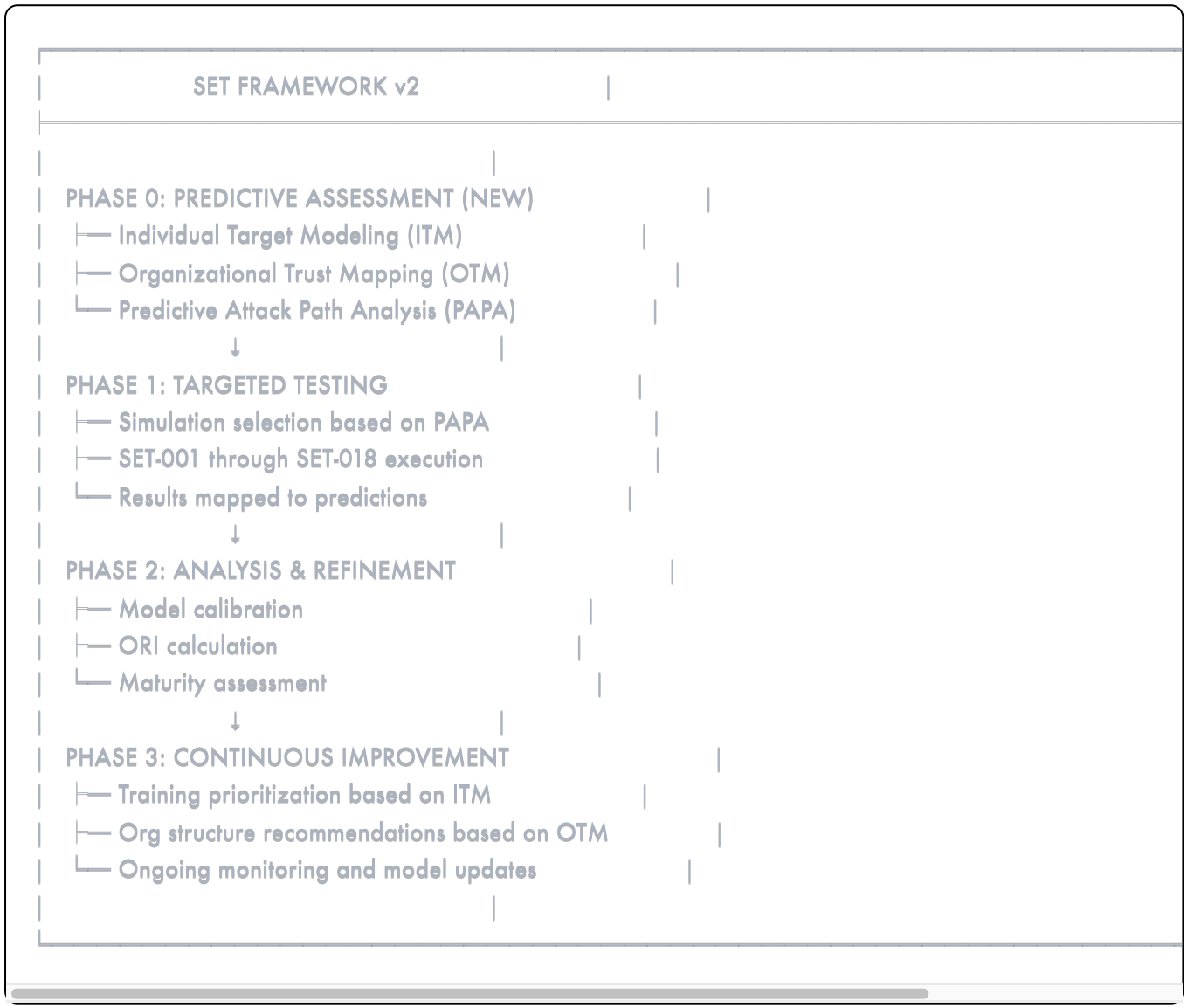
6.3 ORI Maturity Levels

Level	ORI Range	Characteristics
1 - Initial	0.0 - 0.2	No formal program, high vulnerability concentration
2 - Developing	0.2 - 0.4	Basic awareness training, reactive testing

Level	ORI Range	Characteristics
3 - Defined	0.4 - 0.6	PHLRA implemented, targeted testing, metrics tracked
4 - Managed	0.6 - 0.8	Continuous modeling, predictive capabilities, integrated response
5 - Optimizing	0.8 - 1.0	Adaptive resilience, minimal single points of failure, culture of security

7. Integration with SET Framework

PHLRA extends the SET Framework by adding a mandatory pre-assessment phase:



8. Conclusion

The PHLRA extension transforms social engineering testing from reactive measurement to predictive risk management. By modeling individuals, trust relationships, and attack paths *before* testing, organizations gain:

1. **Efficiency**: Test what matters, not everything
2. **Predictability**: Know where you're vulnerable without waiting for an attack

3. **Measurability**: Track ORI over time as a board-level metric
4. **Actionability**: Specific, prioritized remediation based on modeled risk

The human layer is no longer unmeasurable. It simply required the right analytical approach.
