CPF Mathematical Formalization Series - Paper 1: Authority-Based Vulnerabilities: Modelli Matematici e Algoritmi di Rilevamento

Giuseppe Canale, CISSP
Independent Researcher
g.canale@cpf3.org
ORCID: 0009-0007-3263-6897

November 18, 2025

Abstract

Presentiamo la formalizzazione matematica completa degli indicatori della Categoria 1 del Cybersecurity Psychology Framework (CPF): Vulnerabilità Basate sull'Autorità. Ciascuno dei dieci indicatori (1.1-1.10) è rigorosamente definito attraverso funzioni di rilevamento che combinano logica basata su regole, rilevamento di anomalie statistiche e inferenza bayesiana. La formalizzazione consente un'implementazione sistematica in diversi contesti organizzativi mantenendo il fondamento teorico nella ricerca sull'obbedienza di Milgram e nella psicologia sociale contemporanea. Forniamo algoritmi espliciti per il rilevamento in tempo reale, matrici di interdipendenza per l'analisi di correlazione e metriche di validazione per la calibrazione continua. Questo lavoro stabilisce il fondamento matematico per l'operazionalizzazione delle vulnerabilità psicologiche basate sull'autorità nei contesti di cybersecurity.

Keywords: Applied Mathematics, Interdisciplinary Psychology, Computational Statistics, Mathematical Modeling, Cybersecurity Research

1 Introduzione e Contesto CPF

Il Cybersecurity Psychology Framework (CPF) rappresenta un cambio di paradigma dalla consapevolezza di sicurezza reattiva alla valutazione predittiva delle vulnerabilità attraverso la modellazione dello stato psicologico [1]. A differenza dei framework di sicurezza tradizionali che affrontano i controlli tecnici, il CPF identifica sistematicamente le vulnerabilità psicologiche pre-cognitive che creano punti ciechi di sicurezza sistematici.

L'architettura CPF comprende 100 indicatori organizzati in una matrice 10×10, ciascuno fondato su ricerca psicologica consolidata. Il framework impiega un sistema di valutazione ternario (Verde/Giallo/Rosso) mantenendo una rigorosa protezione della privacy attraverso l'analisi comportamentale aggregata piuttosto che la profilazione individuale.

Questa serie di articoli fornisce la formalizzazione matematica completa per ogni categoria CPF, consentendo un'implementazione e validazione rigorose. Ogni indicatore riceve funzioni di rilevamento esplicite, modellazione delle interdipendenze e specifiche algoritmiche. L'approccio matematico serve due scopi: garantire implementazioni riproducibili tra le organizzazioni e stabilire il CPF come metodologia scientificamente rigorosa adatta per la revisione tra pari e la standardizzazione.

La Categoria 1 si concentra sulle vulnerabilità basate sull'autorità, attingendo principalmente dagli studi pionieristici sull'obbedienza di Milgram [2] e dalla successiva ricerca di psicologia sociale sulle dinamiche di autorità nei contesti organizzativi [3]. Queste vulnerabilità sfruttano la tendenza evoluta degli esseri umani a deferire alle figure di autorità percepite, creando debolezze di sicurezza sistematiche che gli attaccanti sfruttano costantemente attraverso campagne di social engineering.

2 Fondamento Teorico: Dinamiche di Autorità

Le vulnerabilità basate sull'autorità emergono dall'intersezione tra psicologia evoluzionistica, cognizione sociale e comportamento organizzativo. Gli esseri umani si sono evoluti in strutture sociali gerarchiche dove la deferenza all'autorità legittima ha migliorato la sopravvivenza [4]. Tuttavia, questi meccanismi adattivi diventano vulnerabilità quando sfruttati da attori malevoli che simulano i marcatori di autorità.

La ricerca dimostra che la conformità all'autorità opera attraverso processi automatici pre-consci [5]. Il riconoscimento dell'autorità si verifica entro 100-200ms dalla presentazione dello stimolo, prima che l'valutazione razionale possa intervenire [6]. Questo vantaggio temporale consente agli attaccanti di aggirare i protocolli di sicurezza consci attraverso la segnalazione rapida dell'autorità.

I modelli matematici qui presentati catturano questi meccanismi psicologici attraverso tre approcci complementari: (1) rilevamento basato su regole per marcatori di autorità espliciti, (2) rilevamento di anomalie per deviazioni statistiche dalle interazioni di autorità di base, e (3) inferenza bayesiana per l'aggiornamento della probabilità basato su fattori contestuali.

3 Formalizzazione Matematica

3.1 Framework di Rilevamento Universale

Ogni indicatore basato sull'autorità impiega la funzione di rilevamento unificata:

$$D_i(t) = w_1 \cdot R_i(t) + w_2 \cdot A_i(t) + w_3 \cdot B_i(t) \tag{1}$$

dove $D_i(t)$ rappresenta il punteggio di rilevamento per l'indicatore i al tempo t, $R_i(t)$ denota il rilevamento basato su regole (binario), $A_i(t)$ rappresenta il punteggio di anomalia (continuo [0,1]), e $B_i(t)$ rappresenta la probabilità posteriore bayesiana. I pesi w_1, w_2, w_3 sommano all'unità e sono calibrati attraverso le baseline organizzative.

L'evoluzione temporale segue lo smoothing esponenziale:

$$T_i(t) = \alpha \cdot D_i(t) + (1 - \alpha) \cdot T_i(t - 1) \tag{2}$$

dove $\alpha = e^{-\Delta t/\tau}$ fornisce il decadimento temporale con costante di tempo specifica dell'organizzazione τ.

3.2 Indicatore 1.1: Conformità Senza Domande

Definizione: Esecuzione automatica di richieste da autorità percepita senza procedure di verifica.

Modello Matematico:

La funzione del tasso di conformità:

$$C_r(t, w) = \frac{\sum_{i \in W(t,w)} E_i}{\sum_{i \in W(t,w)} R_i} \tag{3}$$

dove $W(t, w)$ rappresenta la finestra temporale di larghezza w che termina al tempo t, E_i indica le richieste eseguite, e R_i indica le richieste ricevute dai domini di autorità.

Rilevamento Basato su Regole:

$$R_{1.1}(t) = \begin{cases} 1 & \text{se } C_r(t, 3600) > \theta_{compliance} \\ 0 & \text{altrimenti} \end{cases} \tag{4}$$

dove $\theta_{compliance} = \mu_{baseline} + 2\sigma_{baseline}$ dai dati storici.

Rilevamento di Anomalie: La distanza di Mahalanobis per i pattern di richiesta di autorità multi-variati:

$$A_{1.1}(t) = \sqrt{(\mathbf{x}(t) - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}(t) - \boldsymbol{\mu})} \tag{5}$$

dove $\mathbf{x}(t) = [response\_time, verification\_attempts, escalation\_rate]^T$.

Modello Bayesiano:

$$P(legitimate|factors) = \frac{P(factors|legitimate) \cdot P(legitimate)}{P(factors)} \tag{6}$$

con fattori che includono l'ora del giorno, la reputazione del mittente e i marcatori di urgenza della richiesta.

3.3 Indicatore 1.2: Diffusione di Responsabilità

Definizione: Ridotta responsabilità individuale nelle catene decisionali gerarchiche.

Modello Matematico:

L'indice di diffusione della responsabilità:

$$RD_i(t) = \frac{\sum_{j=1}^{n} T_{ownership}^{(j)}}{n \cdot T_{total}} \tag{7}$$

dove $T_{ownership}^{(j)}$ rappresenta il tempo in cui l'individuo j ha tenuto la responsabilità, e T_{total} è la durata totale dell'incidente.

Funzione di Rilevamento:

$$D_{1.2}(t) = \max\left(0, \frac{N_{transfers}(t) - \mu_{transfers}}{\sigma_{transfers}}\right) \tag{8}$$

dove $N_{transfers}(t)$ conta i trasferimenti di proprietà nel ciclo di vita dell'incidente.

Condizione di Soglia:

$$R_{1.2}(t) = \begin{cases} 1 & \text{se } N_{transfers} > 3 \text{ e } RD_i > 0.7 \\ 0 & \text{altrimenti} \end{cases} \tag{9}$$

3.4 Indicatore 1.3: Suscettibilità all'Impersonificazione di Autorità

Definizione: Vulnerabilità a rivendicazioni di autorità false attraverso canali digitali.

Modello Matematico:

La probabilità di successo dell'impersonificazione:

$$P_{success}(a, c, t) = \sigma(w_a \cdot A(a) + w_c \cdot C(c) + w_t \cdot T(t)) \tag{10}$$

dove σ è la funzione sigmoide, $A(a)$ rappresenta la forza dei marcatori di autorità, $C(c)$ denota la credibilità del canale, e $T(t)$ indica la pressione temporale.

Modello di Correlazione SPF/DKIM:

$$V_{auth}(t) = \frac{\sum_i (1 - SPF_i)(1 - DKIM_i) \cdot Success_i}{\sum_i (1 - SPF_i)(1 - DKIM_i)} \tag{11}$$

Soglia di Rilevamento:

$$R_{1.3}(t) = \begin{cases} 1 & \text{se } V_{auth}(t) > 0.3 \text{ e } N_{failures} > 5 \\ 0 & \text{altrimenti} \end{cases} \tag{12}$$

3.5 Indicatore 1.4: Aggiramento Basato sulla Convenienza

Definizione: Elusione dei controlli di sicurezza per la convenienza percepita dell'autorità.

Modello Matematico:

Il rapporto di aggiramento per convenienza:

$$CBR(t) = \frac{E_{executive}(t)}{E_{standard}(t)} \cdot \frac{T_{standard}}{T_{executive}(t)} \tag{13}$$

dove $E_{executive}$ e $E_{standard}$ rappresentano le concessioni di eccezioni durante le ore esecutive e standard, mentre T rappresenta i periodi di tempo.

Ponderazione Temporale:

$$W(h) = \begin{cases} 1.5 & \text{se } h \in [8, 18] \text{ (orario lavorativo)} \\ 2.0 & \text{se } h \in [18, 22] \text{ (serale esecutivo)} \\ 1.0 & \text{altrimenti} \end{cases} \tag{14}$$

Funzione di Rilevamento:

$$D_{1.4}(t) = CBR(t) \cdot W(hour(t)) \cdot U(urgency(t)) \tag{15}$$

dove $U(urgency)$ pesa gli indicatori di urgenza da 0.5 a 2.0.

3.6 Indicatore 1.5: Conformità Basata sulla Paura

Definizione: Decisioni di sicurezza guidate dalla paura del dispiacere dell'autorità piuttosto che dalla valutazione del rischio.

Modello Matematico:

L'indice di conformità per paura utilizzando l'analisi linguistica:

$$FCI(m) = \sum_i w_i \cdot f_i(m) \tag{16}$$

dove $f_i(m)$ rappresenta la frequenza dei marcatori di paura nel messaggio m, e i pesi w_i sono appresi attraverso l'addestramento supervisionato.

Rilevamento dei Marcatori di Paura: I marcatori di paura includono: {urgent, immediately, critical, must, cannot wait, emergency}

Correlazione del Tempo di Risposta:

$$R_{time}(m) = \frac{T_{response}(m)}{T_{baseline}} \cdot e^{-FCI(m)} \tag{17}$$

Soglia di Rilevamento:

$$R_{1.5}(t) = \begin{cases} 1 & \text{se } FCI > 0.7 \text{ e } R_{time} < 0.3 \\ 0 & \text{altrimenti} \end{cases} \tag{18}$$

3.7 Indicatore 1.6: Effetti del Gradiente di Autorità

Definizione: Inibizione della segnalazione di sicurezza dovuta alla gerarchia organizzativa.

Modello Matematico:

La funzione del gradiente di autorità:

$$AG(i, j) = \frac{H_j - H_i}{H_{max}} \cdot e^{-d(i,j)/\lambda} \tag{19}$$

dove H_i, H_j rappresentano i livelli gerarchici, $d(i,j)$ è la distanza organizzativa, e λ è il parametro di decadimento.

Modello di Inibizione della Segnalazione:

$$P_{report}(i,j) = P_{baseline} \cdot (1 - AG(i,j))^{\beta} \tag{20}$$

dove $\beta > 0$ rappresenta la sensibilità al gradiente di autorità.

Rilevamento Aggregato:

$$D_{1.6}(t) = 1 - \frac{\sum_{i,j} P_{report}(i,j) \cdot I_{incident}(i,j,t)}{\sum_{i,j} I_{incident}(i,j,t)} \tag{21}$$

3.8 Indicatore 1.7: Deferenza all'Autorità Tecnica

Definizione: Accettazione senza domande di rivendicazioni tecniche da esperti percepiti.

Modello Matematico:

Misura della densità di gergo tecnico:

$$TJD(m) = \frac{\sum_{w \in m} I_{technical}(w)}{|m|} \cdot \log\left(1 + \sum_{w \in m} Rarity(w)\right) \tag{22}$$

dove $I_{technical}(w)$ indica il vocabolario tecnico e $Rarity(w)$ misura l'inversione della frequenza delle parole.

Correlazione di Accettazione:

$$P_{accept}(m) = \sigma(\alpha \cdot TJD(m) + \beta \cdot Authority(sender) + \gamma) \tag{23}$$

Rilevamento di Anomalie:

$$A_{1.7}(t) = \frac{TJD(t) - \mu_{domain}}{\sigma_{domain}} \tag{24}$$

dove le baseline specifiche del dominio tengono conto della variazione tecnica legittima.

3.9 Indicatore 1.8: Normalizzazione dell'Eccezione Esecutiva

Definizione: Accettazione graduale degli aggiramento di sicurezza come pratica standard.

Modello Matematico:

La curva di normalizzazione che segue il decadimento della legge di potenza:

$$N(t) = 1 - \left(1 + \frac{t}{t_0}\right)^{-\alpha} \tag{25}$$

dove t_0 rappresenta la costante di tempo e α controlla il tasso di decadimento.

Tracciamento delle Eccezioni Cumulative:

$$E_{cum}(t) = \int_0^t e^{-\lambda(t-\tau)} \cdot E(\tau)\, d\tau \tag{26}$$

con parametro di decadimento esponenziale λ.

Funzione di Rilevamento:

$$D_{1.8}(t) = N(t) \cdot \frac{E_{cum}(t)}{E_{threshold}} \tag{27}$$

3.10 Indicatore 1.9: Prova Sociale Basata sull'Autorità

Definizione: Cascate di conformità innescate da comportamenti approvati dall'autorità.

Modello Matematico:

Il modello di propagazione a cascata:

$$P_{adopt}(i,t) = 1 - \prod_{j \in N(i)} (1 - \alpha_{ij} \cdot A_j(t)) \tag{28}$$

dove $N(i)$ rappresenta il vicinato di rete del nodo i, α_{ij} è il peso dell'influenza, e $A_j(t)$ indica lo stato di adozione.

Amplificazione dell'Autorità:

$$\alpha_{ij} = \alpha_{base} \cdot (1 + \gamma \cdot Authority\_Level(j)) \tag{29}$$

Analisi di Rete: Utilizzando l'analisi degli autovalori laplaciani del grafo per il rilevamento a cascata:

$$\lambda_2 = \min_{x \perp \mathbf{1}} \frac{\mathbf{x}^T L \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \tag{30}$$

3.11 Indicatore 1.10: Escalation di Autorità in Crisi

Definizione: Conformità all'autorità potenziata durante condizioni di crisi percepite.

Modello Matematico:

Fattore di amplificazione della crisi:

$$CAF(t) = 1 + \beta \cdot \tanh(\gamma \cdot Threat\_Level(t)) \tag{31}$$

dove β controlla l'amplificazione massima e γ controlla la sensibilità.

Modello di Conformità Modificato:

$$C_{crisis}(t) = C_{baseline}(t) \cdot CAF(t) \tag{32}$$

Rilevamento di Crisi Multi-fattoriale:

$$Crisis\_Score(t) = \sum_i w_i \cdot f_i(t) \tag{33}$$

con fattori: livello di minaccia esterna, incidenti interni, attenzione dei media e indicatori di stress esecutivo.

4 Matrice di Interdipendenza

Gli indicatori basati sull'autorità mostrano interdipendenze significative catturate attraverso la matrice di correlazione $\mathbf{R}_1$:

$$\mathbf{R}_1 = \begin{pmatrix} 1.00 & 0.65 & 0.45 & 0.55 & 0.70 & 0.40 & 0.35 & 0.60 & 0.50 & 0.75 \\ 0.65 & 1.00 & 0.30 & 0.40 & 0.45 & 0.80 & 0.25 & 0.35 & 0.55 & 0.50 \\ 0.45 & 0.30 & 1.00 & 0.35 & 0.60 & 0.25 & 0.70 & 0.30 & 0.40 & 0.55 \\ 0.55 & 0.40 & 0.35 & 1.00 & 0.50 & 0.30 & 0.25 & 0.75 & 0.45 & 0.40 \\ 0.70 & 0.45 & 0.60 & 0.50 & 1.00 & 0.35 & 0.40 & 0.55 & 0.65 & 0.80 \\ 0.40 & 0.80 & 0.25 & 0.30 & 0.35 & 1.00 & 0.20 & 0.40 & 0.50 & 0.45 \\ 0.35 & 0.25 & 0.70 & 0.25 & 0.40 & 0.20 & 1.00 & 0.30 & 0.35 & 0.40 \\ 0.60 & 0.35 & 0.30 & 0.75 & 0.55 & 0.40 & 0.30 & 1.00 & 0.50 & 0.55 \\ 0.50 & 0.55 & 0.40 & 0.45 & 0.65 & 0.50 & 0.35 & 0.50 & 1.00 & 0.60 \\ 0.75 & 0.50 & 0.55 & 0.40 & 0.80 & 0.45 & 0.40 & 0.55 & 0.60 & 1.00 \end{pmatrix} \tag{34}$$

Le interdipendenze chiave includono:

- Forte correlazione (0.80) tra Conformità Basata sulla Paura (1.5) ed Escalation di Crisi (1.10)
- Alta correlazione (0.75) tra Conformità Senza Domande (1.1) ed Escalation di Crisi (1.10)
- Moderata correlazione (0.75) tra Aggiramento per Convenienza (1.4) e Normalizzazione dell'Eccezione (1.8)
- Significativa correlazione (0.80) tra Diffusione di Responsabilità (1.2) ed Effetti del Gradiente di Autorità (1.6)

5 Algoritmi di Implementazione

Algorithm 1 Valutazione della Vulnerabilità di Autorità

Inizializza i parametri di baseline $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{w}$
for ogni passo temporale t **do**
  Raccogli i dati di telemetria $\mathbf{x}(t)$
  for ogni indicatore $i \in \{1.1, 1.2, \ldots, 1.10\}$ **do**
    Calcola $R_i(t)$ usando la logica basata su regole
    Calcola $A_i(t)$ usando il rilevamento di anomalie
    Calcola $B_i(t)$ usando l'aggiornamento bayesiano
    Calcola $D_i(t) = w_1 R_i(t) + w_2 A_i(t) + w_3 B_i(t)$
    Aggiorna lo stato temporale $T_i(t) = \alpha \cdot D_i(t) + (1 - \alpha) \cdot T_i(t-1)$
  end for
  Calcola le correzioni di interdipendenza usando $\mathbf{R}_1$
  Genera avvisi basati su soglie dinamiche
  Aggiorna le baseline con smoothing esponenziale
  Registra i risultati per la validazione e il rilevamento della deriva
end for

6 Framework di Validazione

Ogni indicatore subisce una validazione continua attraverso molteplici metriche:

Metriche di Classificazione:

$$Precision = \frac{TP}{TP + FP} \tag{35}$$

$$Recall = \frac{TP}{TP + FN} \tag{36}$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{37}$$

Coefficiente di Correlazione di Matthews:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{38}$$

Validazione Temporale: Rilevamento della deriva utilizzando il test di Kolmogorov-Smirnov:

$$D_{KS} = \max_x |F_1(x) - F_2(x)| \tag{39}$$

La ricalibrazione si attiva quando $p < 0.05$.

Protocollo di Validazione Incrociata: La validazione incrociata K-fold con stratificazione temporale garantisce la generalizzazione del modello:

$$CV_{score} = \frac{1}{k} \sum_{i=1}^{k} Performance(Model_i, TestSet_i) \tag{40}$$

7 Conclusione

Questa formalizzazione matematica delle vulnerabilità basate sull'autorità fornisce un fondamento rigoroso per l'implementazione della Categoria 1 del CPF. Ogni indicatore riceve funzioni di rilevamento esplicite che combinano molteplici approcci analitici mantenendo l'efficienza computazionale per l'operazione in tempo reale.

La matrice di interdipendenza cattura importanti correlazioni tra le vulnerabilità relative all'autorità, consentendo un rilevamento potenziato attraverso l'analisi multivariata. Gli algoritmi di implementazione forniscono una guida chiara per l'integrazione del sistema, mentre i framework di validazione garantiscono un'accuratezza sostenuta.

Il lavoro futuro estenderà questo approccio matematico alle restanti nove categorie CPF, creando una specifica formale completa per la valutazione delle vulnerabilità psicologiche nei contesti di cybersecurity. Il rigore matematico consente la ricerca riproducibile, implementazioni standardizzate e validazione obiettiva dell'efficacia del framework CPF.

La categoria delle vulnerabilità basate sull'autorità serve come fondamento per comprendere come le gerarchie organizzative creano punti ciechi di sicurezza sistematici. Formalizzando matematicamente questi meccanismi psicologici, consentiamo il rilevamento e la mitigazione automatizzati delle vulnerabilità che storicamente sono state affrontate solo attraverso programmi soggettivi di consapevolezza della sicurezza.

References

[1] Canale, G. (2024). The Cybersecurity Psychology Framework: A Pre-Cognitive Vulnerability Assessment Model Integrating Psychoanalytic and Cognitive Sciences. *Preprint*.

[2] Milgram, S. (1974). *Obedience to Authority*. Harper & Row.

[3] Zimbardo, P. (2007). *The Lucifer Effect: Understanding How Good People Turn Evil*. Random House.

[4] Henrich, J. (2016). *The Secret of Our Success: How Culture Is Driving Human Evolution*. Princeton University Press.

[5] Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, 54(7), 462-479.

[6] Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, 308(5728), 1623-1626.