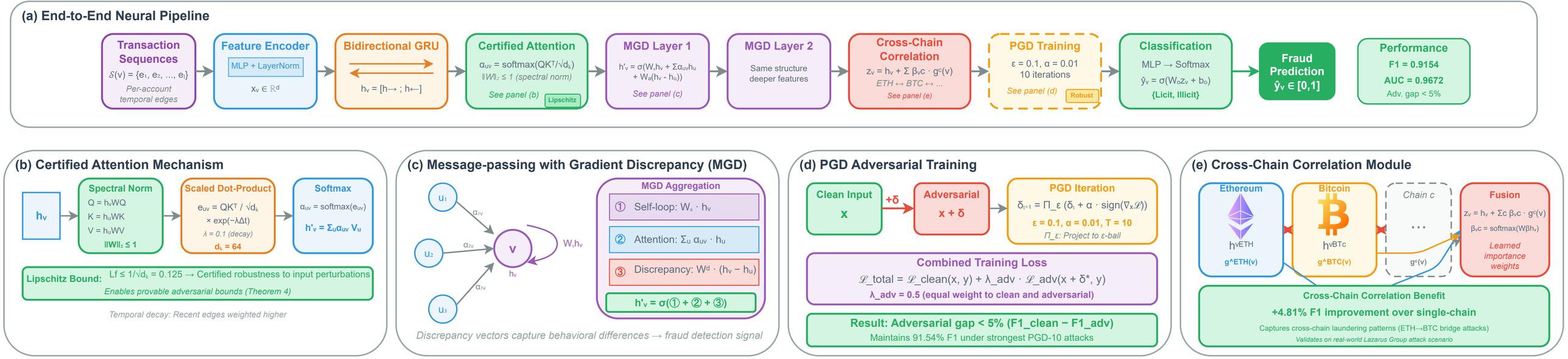
(a) End-to-End Neural Pipeline
Transaction Sequences
S(v) = {e₁, e₂, ..., eⱼ}
Per-account temporal edges
Feature Encoder
MLP + LayerNorm
xᵥ ∈ ℝᵈ
Bidirectional GRU
hᵥ = [h→ ; h←]
Certified Attention
αᵤᵥ = softmax(QKᵀ/√dₖ)
‖W‖₂ ≤ 1 (spectral norm)
See panel (b)
Lipschitz
MGD Layer 1
h'ᵥ = σ(W₁hᵥ + Σαᵤᵥhᵤ + W₂(hᵥ - hᵤ))
See panel (c)
MGD Layer 2
Same structure
deeper features
Cross-Chain Correlation
zᵥ = hᵥ + Σ βᵥc · gᶜ(v)
ETH ↔ BTC ↔ ...
See panel (e)
PGD Training
ε = 0.1, α = 0.01
10 iterations
See panel (d)
Robust
Classification
MLP → Softmax
ŷᵥ = σ(W₀zᵥ + b₀)
{Licit, Illicit}
Fraud Prediction
ŷᵥ ∈ [0,1]
Performance
F1 = 0.9154
AUC = 0.9672
Adv. gap < 5%
(b) Certified Attention Mechanism
hᵥ
Spectral Norm
Q = hᵥWQ
K = hᵤWK
V = hᵤWV
‖W‖₂ ≤ 1
Scaled Dot-Product
eᵤᵥ = QKᵀ / √dₖ
× exp(−λΔt)
λ = 0.1 (decay)
dₖ = 64
Softmax
αᵤᵥ = softmax(eᵤᵥ)
h'ᵥ = Σᵤαᵤᵥ Vᵤ
Lipschitz Bound: Lf ≤ 1/√dₖ = 0.125 → Certified robustness to input perturbations
Enables provable adversarial bounds (Theorem 4)
Temporal decay: Recent edges weighted higher
(c) Message-passing with Gradient Discrepancy (MGD)
u₁
u₂
u₃
v
hᵥ
αᵢᵥ
α₂ᵥ
αₙᵥ
W₁hᵥ
MGD Aggregation
① Self-loop: W₁ · hᵥ
② Attention: Σᵤ αᵤᵥ · hᵤ
③ Discrepancy: Wᵈ · (hᵥ − hᵤ)
h'ᵥ = σ(① + ② + ③)
Discrepancy vectors capture behavioral differences → fraud detection signal
(d) PGD Adversarial Training
Clean Input
x
+δ
Adversarial
x + δ
PGD Iteration
δₜ₊₁ = Π_ε (δₜ + α · sign(∇ₓℒ))
ε = 0.1, α = 0.01, T = 10
Π_ε: Project to ε-ball
Combined Training Loss
ℒ_total = ℒ_clean(x, y) + λ_adv · ℒ_adv(x + δ*, y)
λ_adv = 0.5 (equal weight to clean and adversarial)
Result: Adversarial gap < 5% (F1_clean − F1_adv)
Maintains 91.54% F1 under strongest PGD-10 attacks
(e) Cross-Chain Correlation Module
Ethereum
hᵛETH
g^ETH(v)
Bitcoin
hᵛBTc
g^BTC(v)
Chain c
...
gᶜ(v)
Fusion
zᵥ = hᵥ + Σc βᵥc · gᶜ(v)
βᵥc = softmax(Wβhᵥ)
Learned importance weights
Cross-Chain Correlation Benefit
+4.81% F1 improvement over single-chain
Captures cross-chain laundering patterns (ETH→BTC bridge attacks)
Validates on real-world Lazarus Group attack scenario