SA-BFT Detailed Architecture with Mathematical Formulas

Semantic-Aware Byzantine Fault Tolerance Validator

Input Projection Layer

$$\mathbf{H}_{proj} = \mathbf{z}_{MMF}\mathbf{W}_{proj} + \mathbf{b}_{proj}$$

$$\mathbf{W}_{proj} \in \mathbb{R}^{256 \times 256}, \mathbf{H}_{proj} \in \mathbb{R}^{256}$$

Validator 1

$$\mathbf{Q}_v, \mathbf{K}_v, \mathbf{V}_v = \mathbf{H}_{proj}\mathbf{W}_v^Q, \mathbf{H}_{proj}\mathbf{W}_v^K, \mathbf{H}_{proj}\mathbf{W}_v^V$$

$$\text{Attn}_v(\mathbf{H}) = \text{softmax}\left(\frac{\mathbf{Q}_v\mathbf{K}_v^T}{\sqrt{d_k}}\right)\mathbf{V}_v$$

$$\mathbf{H}_v^{(1)} = \text{Dropout}(\text{Attn}_v(\mathbf{H}_{proj}))$$

$$\mathbf{H}_v^{(2)} = \text{LN}(\mathbf{H}_v^{(1)} + \text{FFN}_v(\mathbf{H}_v^{(1)}))$$

$$\text{FFN}_v(\mathbf{x}) = \max(0, \mathbf{x}\mathbf{W}_{v1} + \mathbf{b}_{v1})\mathbf{W}_{v2} + \mathbf{b}_{v2}$$

$$s_v = \sigma(\text{Linear}(\mathbf{H}_v^{(2)})) \in [0, 1] \text{ (reliability score)}$$

Validator 2

$$\mathbf{Q}_v, \mathbf{K}_v, \mathbf{V}_v = \mathbf{H}_{proj}\mathbf{W}_v^Q, \mathbf{H}_{proj}\mathbf{W}_v^K, \mathbf{H}_{proj}\mathbf{W}_v^V$$

$$\text{Attn}_v(\mathbf{H}) = \text{softmax}\left(\frac{\mathbf{Q}_v\mathbf{K}_v^T}{\sqrt{d_k}}\right)\mathbf{V}_v$$

$$\mathbf{H}_v^{(1)} = \text{Dropout}(\text{Attn}_v(\mathbf{H}_{proj}))$$

$$\mathbf{H}_v^{(2)} = \text{LN}(\mathbf{H}_v^{(1)} + \text{FFN}_v(\mathbf{H}_v^{(1)}))$$

$$\text{FFN}_v(\mathbf{x}) = \max(0, \mathbf{x}\mathbf{W}_{v1} + \mathbf{b}_{v1})\mathbf{W}_{v2} + \mathbf{b}_{v2}$$

$$s_v = \sigma(\text{Linear}(\mathbf{H}_v^{(2)})) \in [0, 1] \text{ (reliability score)}$$

Validator 3

$$\mathbf{Q}_v, \mathbf{K}_v, \mathbf{V}_v = \mathbf{H}_{proj}\mathbf{W}_v^Q, \mathbf{H}_{proj}\mathbf{W}_v^K, \mathbf{H}_{proj}\mathbf{W}_v^V$$

$$\text{Attn}_v(\mathbf{H}) = \text{softmax}\left(\frac{\mathbf{Q}_v\mathbf{K}_v^T}{\sqrt{d_k}}\right)\mathbf{V}_v$$

$$\mathbf{H}_v^{(1)} = \text{Dropout}(\text{Attn}_v(\mathbf{H}_{proj}))$$

$$\mathbf{H}_v^{(2)} = \text{LN}(\mathbf{H}_v^{(1)} + \text{FFN}_v(\mathbf{H}_v^{(1)}))$$

$$\text{FFN}_v(\mathbf{x}) = \max(0, \mathbf{x}\mathbf{W}_{v1} + \mathbf{b}_{v1})\mathbf{W}_{v2} + \mathbf{b}_{v2}$$

$$s_v = \sigma(\text{Linear}(\mathbf{H}_v^{(2)})) \in [0, 1] \text{ (reliability score)}$$

Byzantine Fault Tolerance Consensus Mechanism

$$\mathbf{S} = [s_1, s_2, s_3]^T \text{ (validator reliability scores)}$$

$$\mathbf{W} = \text{softmax}(\mathbf{S}) \text{ (adaptive weights)}$$

$$\text{Consensus}(\mathbf{H}) = \sum_{v=1}^{3} w_v \mathbf{H}_v^{(2)} \text{ (weighted combination)}$$

$$\text{Fault Detection: } f_v = \mathbb{I}[s_v < \tau_{fault}], \tau_{fault} = 0.3$$

$$\text{Byzantine Tolerance: } |\{v : f_v = 1\}| \leq \lfloor \frac{n-1}{3} \rfloor = 1$$

$$\mathbf{H}_{validated} = \text{Consensus}(\mathbf{H}) \text{ if valid, Fallback}(\mathbf{H}_{proj}) \text{ if Byzantine}$$

Validated Features Output

$$\mathbf{z}_{SA-BFT} = \mathbf{H}_{validated} \in \mathbb{R}^{256}$$

Fault-tolerant, semantically-aware validated features

$$\text{Reliability}: R = \frac{\sum_v s_v \cdot (1 - f_v)}{\sum_v (1 - f_v)} \in [0, 1]$$

SA-BFT Mathematical Specifications:

- Parameters: $\theta_{SA-BFT} = 2.1M$ (13.4%)
- Input Dimension: $d_{in} = 256$
- Validators: $n_v = 3$ (parallel)
- Attention Heads: $h_v = 8$ per validator
- Dropout Rate: $p_{drop} = 0.15$
- Fault Threshold: $\tau_{fault} = 0.3$
- Byzantine Tolerance: $f \leq 1$ (max faults)
- Consensus Weight: $w_v \propto s_v$
- Response Time: $T_{response} < 10ms$
- Detection Accuracy: $A_{fault} > 95\%$
- Output Dimension: $d_{out} = 256$
- Reliability Score: $R \in [0, 1]$