[CAPABILITY PROFILE · AUTHREX-04]

MAIVA · MULTI-AGENT INTEGRITY VOTING

Aggregate authority across multiple agents so that a single compromised node, or a minority acting in concert, cannot carry the outcome, while treating agreement as evidence rather than truth.

MODELED **TESTED** NOT YET VALIDATED

Name variants: Deposit and patents-page title: Multi-Agent Integrity Verification. Same framework; MIT-licensed deposit v5.18.

PRIMARY FUNCTION	Byzantine-resilient aggregation using a trimmed weighted median resistant to f adversaries in a 3f+1 roster, with CUSUM-augmented anomaly detection and an action-gate classification informed by DoDD 3000.09 categories (author mapping).
MISSION AREAS	Swarm and multi-platform operations; distributed sensing; federated model contributions; any decision aggregated across heterogeneous actors.
EVIDENCE ANCHORS	Zenodo DOI 10.5281/zenodo.19015517 (MIT); TLA+ specification; 37 self-tests; faithful model MAIVA_QUORUM.tla; seeded browser simulation; BLADE-SWARM refinement (5 safety, 3 liveness, model checked at stated bounds).
CURRENT STANDING	Demonstrated in simulation; one faithful formal model located a documented open defect in production source (quorum at n = 1); the repair is specified, not shipped; not independently reproduced.
SELF-ASSESSED TRL	TRL 3 for the aggregation logic; the quorum defect is open. No formal TRL determination has been performed.
INTELLECTUAL PROPERTY	U.S. provisional application 64/110,221, filed July 13, 2026 (receipt 78659350), reported as filed; unexamined; confers no enforceable rights. Deposit under MIT license.

[MISSION PROBLEM]

In multi-agent operations a single compromised agent, or a majority contaminated by the same false information, can carry the outcome. Consensus is not truth: majority voting assumes independent failure, and a quorum rule whose threshold is derived from the roster it is checking can only report on itself. The mission consequence is a swarm that commits to an action because its own bookkeeping says enough members agreed, when the agreement was manufactured.

[OPERATIONAL RELEVANCE]

- Attritable swarm autonomy (BLADE-SWARM reference design instantiates MAIVA with sub-quorum decomposition).
- Distributed sensing and multi-platform track fusion in contested communications.
- Federated learning contributions where a poisoned update must not be admitted on the strength of agreement alone (AUTHREX-DA extends this to data).
- Redundant computing where dissimilar channels must be weighed, not merely counted.

[CAPABILITY DESCRIPTION]

MAIVA aggregates authority-relevant evidence across agents before the tier is issued. Each agent runs local SATA and HMAA evaluation and contributes an attested vote to a Mission Authority Node, where trimmed weighted median aggregation discards the f most extreme contributions in a 3f+1 roster and weights the remainder by trust; a three-layer CUSUM detector watches for slow drift in any member, and graduated escalation assigns per-level action permissions. The output is an integrity-weighted input to HMAA, not a decision. The faithful model MAIVA_QUORUM located a defect in the shipped quorum check: the fault bound f was derived from the roster size and the roster size was then compared against a threshold built from that bound, so at n = 1 a single agent constitutes a quorum. The repair requires the maximum compromised count to be an independent adversary assumption (MaxCompromised) stated beside the requirement, and the record notes that no runtime check can establish that bound: a system cannot verify its own compromise level.

[MAIVA · ARCHITECTURE, ASSURANCE, EVIDENCE]

[TECHNICAL ARCHITECTURE]

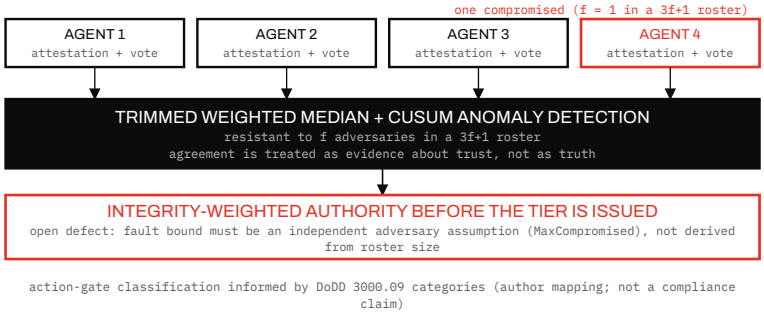

Figure 1. MAIVA technical architecture. Encoding: solid = implemented; dashed = specified; gray = external; filled = enforcement point.

[HOW IT WORKS]

1. Collect attested contributions from every agent in the roster with their current SATA trust.
2. Apply the fault-tolerance precondition: the number of compromised members must not exceed the declared f (an assumption, not a measurement).
3. Trim the f most extreme contributions; compute the trust-weighted median of the remainder.
4. Run CUSUM anomaly detection per member to flag drift below the trim threshold.
5. Classify the pending action by gate category (author mapping to DoDD 3000.09 categories) to decide whether integrity-weighted authority may proceed or must escalate.
6. Emit the integrity-weighted authority input to HMAA; flag members whose contributions were trimmed or whose CUSUM fired.
7. Record the roster, weights, trimmed members, and the declared f to the ledger.

[TECHNICAL DIFFERENTIATION]

Triple-modular redundancy and N-version voting mask a faulty channel by majority agreement and assume independent failure; common-mode conditions such as spoofing defeat them because every channel agrees and every channel is wrong. MAIVA weights agreement by trust and treats it as evidence about trust rather than as truth. Its own published defect shows the reverse risk: a quorum rule that is self-referential reports on itself, which is why the fault bound must be an external assumption.

[ASSURANCE MECHANISMS: DESIGN PROPERTY VERSUS ESTABLISHED PROPERTY]

Mechanism	Status
Trimmed weighted median tolerates f adversaries in 3f+1	Design property; exercised in simulation and self-tests
Q_NEVER_FAILS over the shipped quorum check	Held across 12,507,501 distinct states, in the narrowed sense that the self-referential check is satisfied for every roster size 1 to 5,000, which is the defect: at n = 1 one agent is a quorum
Repaired quorum (Quorum + MaxCompromised)	Holds at f = 1 and f = 2 in the repaired model; not shipped in the production engine
R1_ASSUMPTION: compromise bound is external	Stated requirement; no runtime check can establish it
BLADE-SWARM refinement S1 to S5, L1 to L3	Model checked at stated bounds (5 safety invariants, 3 liveness properties)
Strong form: attester compromise inside a met quorum	FAILED in the naive-versus-strong comparison; repaired by Quorum + MaxCompromised

[TECHNICAL EVIDENCE]

Evidence	Result	Source / artifact
Faithful model of the quorum decision	MAIVA_QUORUM.tla FAITHFUL; defect located at named source lines	authrex.systems/formal-coverage
Quorum-check state space	12,507,501 distinct states; check satisfied for n = 1 to 5,000 (the defect)	formal-coverage record
Self-tests	37 self-tests in the MIT deposit v5.18	Zenodo 10.5281/zenodo.19015517
Seeded browser simulation	Byzantine minority scenarios; deterministic replay	burakoktenli.com/maiva-simulation
BLADE-SWARM TLA+ refinement	5 safety, 3 liveness at stated bounds; N = 10, 50, 500 in simulation	Zenodo 10.5281/zenodo.20351198

[EVIDENCE SNAPSHOT]

EVIDENCE SNAPSHOT	
Implemented	Yes (MIT deposit)
Executable artifact	Yes
Formal model	Yes (one faithful)
Simulation evidence	Yes
Hardware validation	No
Field validation	No
Independent validation	No
Government evaluation	No
Operational deployment	No

[LIMITATIONS]

- Documented open defect: the shipped quorum check is self-referential; the repair exists as a specification, not as released code.
- Multi-agent evaluation uses simulated agent populations; no physical multi-agent experiment has been performed.
- The fault bound f is an assumption that no runtime check can establish; correlated compromise beyond f is outside the guarantee.
- MAIVA.tla and MAIVA_REPAIRED.tla are surrogate countermodels: findings about the mechanism class, not about the implemented code.
- Simulation and formal-model evidence only; all data synthetic; no hardware-in-the-loop test, no field validation, no independent replication, no deployment.
- Not independently reproduced or technically verified by another party; second-environment reruns are confirmation by the same producer.

[TECHNOLOGY READINESS (SELF-ASSESSED; NO FORMAL TRA)]

CURRENT ASSESSED TRL	TRL 3 (self-assessed against the DoD nine-level scale; no formal technology readiness assessment has been performed).
BASIS	Aggregation logic implemented and exercised in simulation with a formal model of one critical decision; the located defect means the shipped artifact does not yet satisfy its own stated requirement.
EVIDENCE SUPPORTING	concept and specification; faithful model of the quorum decision; self-tested implementation; simulation demonstration
EVIDENCE REQUIRED FOR NEXT TRL	(1) Ship the Quorum + MaxCompromised repair and re-check with the faithful model. (2) Multi-agent software-in-the-loop with independent agent processes and injected compromise (Phase 4). (3) Independent replication of the 37 self-tests and the TLC configurations.

[INTEGRATION PROFILE]

INPUTS	Attested contributions per agent; per-agent trust; declared fault bound f (MaxCompromised) from mission policy.
OUTPUTS	Integrity-weighted authority input; trimmed-member and CUSUM flags; ledger entry.
INTERFACES	Agent-to-aggregator message channel; HMAA input; BLADE-SWARM ICD-SWARM-001 (design).
DEPLOYMENT	Software module; requires more than one modality or actor to be meaningful.
SECURITY BOUNDARY	Inside: aggregation, CUSUM, roster bookkeeping. Outside: agents, their attestation roots, the channel.
DATA REQUIREMENTS	Synthetic; multi-platform recorded telemetry would enable Phase 2.

[TRANSITION ROADMAP]

Phase	Activity for this capability	Who can execute it	Exit evidence
1 Technical review	Review of the artifacts listed in the evidence table	Program alone, with a government reviewer	Reviewer findings; correction log entries
2 Government-defined simulation	Sponsor scenarios on the seeded console	Program with sponsor scenarios	Deterministic, replayable results; hash-chained records
3 Degradation and failure injection	Evidence lapse, spoofing, delayed human response	Program with sponsor parameters	False-restriction, miss, and contraction-latency figures
4 Government testbed / SIL	Integration with a representative controller or interface	Requires sponsor interface and testbed	SIL integration report
5 Hardware-in-the-loop	Representative interference; timing on real hardware	Requires sponsor laboratory or rig	Measured latencies; TRL 4 to 5 evidence
6 Relevant-environment demonstration	Independent evaluation and red team	Requires independent evaluator	Independent report
7 Program integration	Only after earlier gates succeed	Requires transition partner	Not claimed to be available

[GOVERNMENT ENGAGEMENT REQUEST (SPECIFIC)]

- Adversarial evaluation of multi-agent agreement by a swarm or multi-platform test organization.
- A government-stated assumption on tolerable compromise (f) for the target mission, which the architecture cannot derive.
- Access to a multi-vehicle software-in-the-loop environment for Phase 4.

[REVIEWER QUESTIONS THIS PROFILE ANSWERS]

What is it: see capability description. Why does DoD care: see operational relevance. Where would it fit: see integration profile. Does it exist: see evidence snapshot. What evidence supports it: see technical evidence. Has anyone independent tested it: no. Has it been tested in the real world: no. What does it rely on: see limitations and security boundary. What can cause it to fail and what does it do then: see assurance mechanisms and the published counterexamples. How mature: see the TRL block. What would the government need to provide: see the request above. Next measurable milestone: first item under evidence required for next TRL.