Internet Engineering Task Force
Internet-Draft
Intended status: Experimental
Expires: 5 December 2026

L. Melegassi
Catellix

3 June 2026

Botnet Identification by Coordination-Coherence and Coherence-Driven Remediation Signaling: The MVPS Botnet Profile draft-melegassi-mvps-botnet-coherence-00

Abstract

This document specifies how the Multi-Vantage Path Synchrony (MVPS) framework [I-D.melegassi-ippm-mvps-bundle] and its DDoS profile [I-D.melegassi-mvps-ddos-resilience] are extended to IDENTIFY the participating sources of a botnet by their coordination-coherence signature, and to EMIT corroborated, signed evidence that DRIVES existing, standardised remediation ("sanitization") machinery.

The central design constraint is honesty about scope. MVPS does NOT itself clean, quarantine, sinkhole, or take down infected hosts. Remediation is performed by the mechanisms already defined by the IETF:

- RFC 6561 (Recommendations for the Remediation of Bots in ISP Networks) -- the notification/remediation workflow;
- RFC 9132 / RFC 8783 / RFC 8811 (DOTS) -- mitigation request signaling;
- RFC 8520 (Manufacturer Usage Description, MUD) -- containment of compromised constrained/IoT devices;
- BCP 38 / BCP 84 (RFC 2827 / RFC 3704) -- source-address validation against spoofed botnet traffic;
- RFC 7970 / RFC 8727 (IODEF) and RFC 6545 (RID) -- the exchange and inter-domain coordination formats;
- RFC 9424 -- the Indicator-of-Compromise (IoC) framing for what MVPS exports.

What MVPS contributes is precisely the gap RFC 6561 Section 4 names: it asks operators to "confirm a bot infection through the use of a combination of multiple bot detection data points ... to corroborate information of varying dependability ... [and] avoid or minimize the possibility of false-positive identification of hosts." MVPS is exactly such a corroboration engine, with the addition of a provable false-positive bound (Theorem B2) and a coordination-coherence test (Theorem B1) that distinguishes a genuinely coordinated population (a botnet) from an equal number of independently misbehaving hosts.

We state three results:

Theorem B1 (Coordination Signature). A population of S sources driven by a common controller produces a low-rank deformation of the cross-vantage coherence covariance; independent legitimate sources do not. The leading eigenvalue ratio is therefore a detector of coordination, not of volume.

Theorem B2 (Corroboration / False-Positive Bound). If a single vantage flags a candidate source with per-vantage false-positive rate p, then requiring agreement across V independent vantages drives the host-level false-positive probability to at most p^V

under vantage independence, and to a stated mixture bound under partial correlation.

Theorem B3 (No Unilateral Action / Remediation Soundness). MVPS emits evidence only. Every enforcement step is taken by an existing standardised control point (RFC 6561 / DOTS / MUD / BCP 38). No host is quarantined on single-vantage evidence.

Theorem B4 (Falsifiability / coherence-collapse axis). The corroboration bound of B2 COLLAPSES on a correlated benign population: a legitimate flash crowd is coordinated-but-benign, the botnet analogue of the COHERENT_BUT_FALSE failure mode of the MVPS AI-Coherence extension [I-D.melegassi-mvps-ai-coherence]. When the coherence environment so collapses, that extension's falsifiability axis enters: re-test the apparent coordination on the machine-regularity subspace -- features a human crowd cannot fake. A flash crowd collapses to the independent floor there; a real bot fleet does not.

Theorem B5 (No Free Decorrelation). Spreading the botnet's coordination across many sources to drop each per-vantage signal cannot lower what the multi-vantage aggregate sees: the coherent statistic is spread-INVARIANT (T_agg = sqrt(E)) with NO compute term, so the multi-vantage advantage GROWS with the spread and the silent-coordination cap is E < tau^2. This is the exact form of the B1 evasion corollary.

Theorem B6 (Non-Blinding of the corroboration set). Silently hiding the coordination by corrupting the vantages is impossible while the redundancy rho = V - d_eff >= 1 with diverse vantages: any such blinding needs k > rho corruptions and is FLAGGED by the vantage-integrity monitor (a non-zero stealth-gap), and the only un-flagged corruption -- forging vantage reports -- is gated by a post-quantum signature (ML-DSA, FIPS 204). "Blind" implies "known-blind".

THE THESIS IN ONE LINE. Cross-vantage agreement is necessary but not sufficient: the coherence environment can collapse (correlated benign crowds, or Byzantine vantages), and where it collapses the AI-coherence axes -- falsifiability (B4) and Byzantine-robust geometric-median aggregation [I-D.melegassi-mvps-ai-coherence] -- are what keep the identification sound.

NOTE ON DATA PROVENANCE. Section 7 reports two kinds of result, each tagged. Section 7.1 is a LABELLED SYNTHETIC ground-truth experiment (script scripts/simulate_botnet_coherence.py). Sections 7.2 and 7.3 are measured on REAL labelled botnet traffic: the CTU-13 dataset of the Stratosphere IPS Laboratory (bidirectional NetFlow [RFC5103] / IPFIX [RFC7011] records labelled Botnet / Normal / Background), across three malware families (Neris, Rbot, Virut). On that real data the detector separates botnet from normal traffic with held-out AUC 0.85-0.999, and the multi-vantage advantage (Theorem B5) is instantiated with the MEASURED per-flow effect size. What remains REQUIRED future work (Section 10) is corroboration across THREE OR MORE INDEPENDENT REAL VANTAGES observing the same event (the real-data form of Theorem B2): CTU-13 is a single capture point. No claim of operational botnet takedown is made.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute

working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 5 December 2026.

Copyright Notice

Table of Contents

1. Introduction 3
   1.1. The honest distinction: identify vs sanitize 3
   1.2. Relationship to the MVPS DDoS profile 4
   1.3. Conventions used in this document 4
2. The RFC Landscape This Profile Plugs Into 5
   2.1. RFC 6561 -- the remediation workflow 5
   2.2. DOTS -- mitigation request signaling 5
   2.3. MUD -- constrained-device containment 6
   2.4. BCP 38 / BCP 84 -- anti-spoofing 6
   2.5. IoC / IODEF / RID -- evidence exchange 6
   2.6. Vantage integrity: RPKI/ROV, SAVI, and PQC identity 6
   2.7. Real-world grounding (documented incidents / CVEs) 7
3. Scope and Threat Model 7
4. The Coordination-Coherence Signature 7
5. Theorems and Proofs 8
   5.1. Theorem B1: Coordination Signature 8
   5.2. Theorem B2: Corroboration / False-Positive Bound 9
   5.3. Theorem B3: No Unilateral Action 10
   5.4. Theorem B4: Falsifiability / coherence-collapse axis 10
   5.5. When vantages collapse: Byzantine-robust aggregation 10
   5.6. Theorem B5: No Free Decorrelation (multi-vantage) 10
   5.7. Theorem B6: Non-Blinding of the corroboration set 10
6. From Evidence to Sanitization (the pipeline) 10
   6.1. Evidence object (IoC, RFC 9424 framing) 10
   6.2. Hand-off to RFC 6561 remediation 11
   6.3. Hand-off to DOTS / MUD / BCP 38 11
   6.4. Canonical export: YANG and JSON 11
7. Results: detection on labelled ground truth (synthetic+real) ..12
   7.1. Synthetic labelled ground truth 12
   7.2. Real labelled botnet traffic (CTU-13) -- detection 13
   7.3. The multi-vantage advantage on real effect sizes 14
8. Security Considerations 15
9. Privacy Considerations 16
10. Operational and Validation Considerations 16
11. IANA Considerations 17
12. References 17

Appendix A. Reproducibility (validators, simulations, receipts) ..19
Appendix B. Implementation and Deployment Guidance 21
Acknowledgements 22
Author's Address 22

1. Introduction

The MVPS DDoS profile [I-D.melegassi-mvps-ddos-resilience] proves that a volumetric or distributed attack is DETECTED and the hit region ATTRIBUTED in time (M-1)*T_tick, independent of attack volume. That profile answers "is there an attack, and where is it landing?" It does not answer "WHICH sources are participating, are they a coordinated botnet, and what corroborated evidence can be handed to a remediation process?"

This document answers the second question. It treats the botnet problem as two strictly separated phases:

(a) IDENTIFICATION -- recognising, with a bounded false-positive rate, that a set of sources is behaving as one coordinated population; and

(b) SANITIZATION -- the operational remediation of those sources, which this document deliberately delegates, in full, to existing IETF mechanisms.

The contribution is confined to phase (a) plus the clean hand-off to phase (b).

1.1. The honest distinction: identify vs sanitize

It is tempting to claim that a detector "sanitizes" a botnet. This document does not make that claim and actively guards against it. "Sanitization" -- notification of the subscriber, walled-garden quarantine, sinkholing of command-and-control (C2), device containment, or upstream scrubbing -- changes the state of a third party's host or traffic. Such action carries legal, privacy, and collateral-damage risk and is, by long-standing IETF consensus (RFC 6561), the province of the network operator under defined process, not of a monitoring instrument.

What a monitoring instrument can legitimately do is reduce the uncertainty that makes remediation risky. RFC 6561 Section 4 is explicit that the hard part of bot remediation is corroboration: confirming infection from multiple independent data points to "avoid or minimize the possibility of false-positive identification of hosts." MVPS is designed to be exactly that corroborating data source, with the property -- not present in single-sensor pipelines -- that its false-positive rate at the host level is bounded in closed form by the number of independent vantages that agree (Theorem B2).

1.2. Relationship to the MVPS DDoS profile

This profile REUSES, without modification, the canonical machinery: the per-vantage coherence vector x_v(t) in R^d and the Mahalanobis D^2 statistic with chi-square phase thresholds chi^2_{d,0.95} / chi^2_{d,0.99} [I-D.melegassi-mvps-incremental-be], the cell partition and cell-aware minimax aggregation D^2_minimax over k cells with Byzantine bound floor((k-1)/2) [I-D.melegassi-mvps-ddos-resilience], the M-multiplier / T_tick detection cadence, and the sub-tick transport of [I-D.melegassi-coherence-bfd]. In particular, per-source coherence data rides the existing Coherence-BFD TLVs -- the Vantage-Sketch TLV

(type 0xE0) and the AuthHMAC-SHA256 TLV (type 0xE9) -- with no new wire format. The only new machinery is:

- a per-source (rather than per-cell) coherence projection;
- the leading-eigenvalue coordination test (Theorem B1);
- the V-vantage corroboration rule (Theorem B2);
- the evidence-export and hand-off pipeline of Section 6, including the YANG module and JSON schema of Section 6.4; and
- no actuation: evidence only (Theorem B3).

No new wire format and no new cryptographic primitive are introduced; authentication (the 0xE9 AuthHMAC-SHA256 TLV), replay protection (monotonic BFD sequence numbers), and control-plane isolation are inherited from the referenced documents.

1.3. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

"Vantage" and "broker" are as defined in [I-D.melegassi-mvps-ddos-resilience]. "Source" denotes an external IP address (or, where SAVI/BCP 38 applies, a validated source) seen by one or more vantages. "Coordination" denotes statistically shared timing/behaviour across sources that exceeds what independent legitimate sources produce. "Sanitization" denotes any remediation action defined by RFC 6561 and is OUT OF SCOPE for the detector itself.

2. The RFC Landscape This Profile Plugs Into

This profile is intentionally a thin layer of NEW analysis on top of a mature set of EXISTING standards. Each existing mechanism owns a phase that MVPS must not duplicate. The overarching framing of the threat is the one set out in the IAB’s Internet Denial-of-Service Considerations [RFC4732]; this profile adds corroborated detection evidence within that frame and defers every mitigation phase to the standards below.

2.1. RFC 6561 -- the remediation workflow

RFC 6561 defines the ISP-side bot remediation lifecycle: detection, notification, remediation, and failure handling, with strong privacy and non-disruption requirements. Its Section 4 calls for multi-point corroboration to avoid false-positive host identification, and its Section 5 governs subscriber notification.

MVPS placement: MVPS is a detection/corroboration input to the RFC 6561 process. It MUST NOT trigger notification or quarantine directly; it produces the corroborated evidence on which an RFC 6561 process MAY act.

2.2. DOTS -- mitigation request signaling

DOTS [RFC9132] (signal channel, obsoleting RFC 8782), [RFC8783] (data channel), and [RFC8811] (architecture) let a client request upstream DDoS mitigation.

MVPS placement: when the identified coordinated population is actively flooding a protected resource, an MVPS broker MAY act as a DOTS client and request mitigation, carrying the per-source evidence

set as DOTS aliases / scope. The decision to request mitigation is an operator policy, not an automatic consequence of detection.

2.3. MUD -- constrained-device containment

MUD [RFC8520] lets a device's manufacturer publish the device's intended communication profile so the network can confine it.

MVPS placement: for the IoT botnet class (e.g., Mirai-style), MVPS-identified compromised devices that possess a MUD profile can be contained by re-asserting that profile at the access network. MVPS supplies the "this device is now deviating from its MUD" signal with cross-vantage corroboration; enforcement is MUD's.

2.4. BCP 38 / BCP 84 -- anti-spoofing

BCP 38 [RFC2827] and BCP 84 [RFC3704] specify source-address validation (ingress filtering). Botnet traffic frequently spoofs source addresses; a spoofed source cannot be remediated by host notification.

MVPS placement: MVPS evidence is host-actionable only for sources that survive source-address validation. This profile therefore REQUIRES that per-source evidence be tagged with whether the source passed SAVI/BCP 38 validation; un-validated sources are reported as "spoof-suspect" and routed to traffic-level mitigation (DOTS), not to host-level remediation (RFC 6561).

2.5. IoC / IODEF / RID -- evidence exchange

RFC 9424 frames what an MVPS finding IS: a network-level Indicator of Compromise (an IP/prefix/behavioural artefact), positioned on the RFC 9424 "pyramid of pain" at the network-indicator tier. IODEF [RFC7970] and its JSON binding [RFC8727] are the document formats for sharing such findings; RID [RFC6545] is the inter-domain request/coordination transport.

MVPS placement: MVPS exports each corroborated finding as an IoC carried in an IODEF document, with the coherence statistics and the V-vantage agreement count attached as confidence metadata.

2.6. Vantage integrity: RPKI/ROV, SAVI, and PQC identity

Theorems B2-B6 are only as sound as the vantages themselves. Three existing mechanisms -- none invented here -- are what keep the redundancy margin rho >= 1 of Theorem B6 real and the BCP 38 tag of Section 2.4 meaningful at host granularity:

- RPKI [RFC6480] and BGP prefix origin validation [RFC6811], distributed by the RPKI-to-Router protocol [RFC8210], protect the route-view class of vantage against the BGP-hijack poisoning that the AI-Coherence cascade model ([I-D.melegassi-mvps-ai-coherence] Section 15) quantifies: a hijack that would silently move a vantage's view is rejected (Invalid) rather than accepted, preserving vantage diversity (rho) instead of collapsing it.

- SAVI [RFC7039] realises BCP 38 / BCP 84 at per-host binding granularity, so the bcp38_validated tag of Section 2.4 / 6.1 reflects an actual source-binding state, not a coarse prefix assumption. Without SAVI a "validated" tag is only as good as the nearest ingress filter.

- PQC vantage identity: Theorem B6(iii) requires each vantage report to be unforgeable against a quantum adversary. This profile RECOMMENDS binding each report with a NIST

post-quantum signature, ML-DSA [FIPS204], hardware-rooted, either replacing or wrapping the inherited Coherence-BFD AuthHMAC-SHA256 TLV (0xE9).

2.7. Real-world grounding (documented incidents / CVEs)

The mechanisms above are not hypothetical. This subsection is INFORMATIVE: the CVEs motivate the threat model and the hand-off targets; they are not used as proof of any theorem (the proofs are in Section 5 and the appendix).

- IoT command-and-control botnets (Mirai class) recruited hosts through device vulnerabilities such as CVE-2017-17215 (Huawei HG532), CVE-2014-8361 (Realtek SDK miniigd), and CVE-2018-10561 / CVE-2018-10562 (Dasan/GPON). These are exactly the constrained-device population that the MUD [RFC8520] hand-off (Section 6.3) is designed to contain, and their shared C2 is the shared command direction g of Theorem B1.

- Reflection/amplification floods exploit exposed UDP services, e.g. CVE-2018-1000115 (memcached UDP, the 1.35 Tbps class event). Such volumetric, often spoofed, coordinated floods are the DOTS [RFC9132] traffic-mitigation and BCP 38 spoof-suspect path (Sections 2.2, 2.4, 6.3), and illustrate why Theorem B5’s volume/spread independence matters.

- Mass-exploitation events such as CVE-2021-44228 (Log4Shell) drove simultaneous, identically-templated requests from many sources -- a textbook coordinated command with a strong shared direction g, i.e. the high-lambda_ratio signature of Theorem B1.

In each case MVPS would IDENTIFY the coordinated population and EMIT corroborated evidence; remediation remains owned by RFC 6561 / DOTS / MUD / BCP 38 (Theorem B3).

3. Scope and Threat Model

In scope:

- Identifying that N_obs observed sources contain a coordinated sub-population of size S (Theorem B1).
- Bounding the false-positive probability of labelling any individual source as a member (Theorem B2).
- Exporting corroborated evidence to RFC 6561 / DOTS / MUD / IODEF (Section 6).

Out of scope (delegated or future):

- Any host- or traffic-state change ("sanitization") -- owned by RFC 6561 / DOTS / MUD / BCP 38.
- Malware classification, C2 protocol reverse-engineering, or attribution to a threat actor.
- Operation against an adversary who controls a strict majority of vantages or cells (Byzantine bound inherited from [I-D.melegassi-mvps-ddos-resilience] Theorem D2).

Adversary model:

A1. Coordinated population. S sources receive correlated commands (timing, target, payload shape) from one or more controllers. This correlation is the signal.

A2. Decorrelation evasion. The adversary jitters or spreads per-source behaviour to suppress the coordination signature; this is bounded in Section 5.1 (a cost, not a free evasion) and made EXACT in Section 5.6 (Theorem B5): the coherent coordinated effect seen by the multi-vantage aggregate is spread-invariant, so spreading thins the per-vantage signal but never the aggregate.

A3. Spoofing. Sources spoof addresses; handled by the BCP 38 tagging requirement of Section 2.4 (spoofed sources cannot be host-remediated and are routed to traffic mitigation).

A4. Vantage corruption / blinding. The adversary corrupts or forges vantages to make the coordination invisible. Bounded in Section 5.7 (Theorem B6): silent blinding is impossible while redundancy rho >= 1 with diverse vantages, and the only un-flagged path is PQC-gated forgery (Section 2.6).

4. The Coordination-Coherence Signature

Each vantage v, each tick t, observes a set of sources and computes a per-source feature vector f_{v,s}(t) in R^d (e.g., arrival-rate, inter-arrival regularity, destination entropy, flag mix, TTL stability). Stacking sources gives a matrix F_v(t).

The key empirical premise (made falsifiable in Section 10): a COORDINATED population's feature matrix is approximately LOW RANK, because many sources move together. An equally large set of INDEPENDENT legitimate sources yields a near-full-rank, near-diagonal feature covariance.

Define the cross-source coherence covariance at vantage v:

$$C_v(t) = (1/n) * F_v(t)^T F_v(t) \quad \text{(centred)}$$

and the leading eigenvalue ratio:

$$\text{lambda\_ratio}_v(t) = \lambda_1(C_v(t)) / \text{trace}(C_v(t)).$$

A high lambda_ratio indicates energy concentrated in one direction -- the coordination signature. This is the per-source analogue of the per-cell D^2 used for DDoS detection.

5. Theorems and Proofs

5.1. Theorem B1 (Coordination Signature)

STATEMENT. Let S sources be driven by a common controller such that each source's feature vector is f_s = a_s * g + e_s, where g is a shared command direction, a_s a per-source gain, and e_s independent zero-mean noise with per-coordinate variance sigma^2. Let an equal number of independent legitimate sources have f_s = e'_s with e'_s independent zero-mean, variance sigma^2. Then, as S grows, the expected leading eigenvalue ratio of the coordinated population is bounded below by

$$E[\text{lambda\_ratio\_coord}] \geq (||g||^2 * \text{Avar}) / (||g||^2 * \text{Avar} + d*\sigma^2)$$

where Avar = E[a_s^2], while for the independent population

$$E[\text{lambda\_ratio\_indep}] \to 1/d \quad \text{as } S \to \infty.$$

PROOF (sketch). For the coordinated population, C = ||g||^2 * Avar * (gg^T/||g||^2) + sigma^2 * I + o(1) by the law of large numbers over S, so lambda_1 -> ||g||^2*Avar + sigma^2 and trace -> ||g||^2*Avar + d*sigma^2, giving the stated ratio. For the independent population C -> sigma^2 * I, whose eigenvalues are equal, so lambda_1/trace -> 1/d. The two regimes are separated by a gap that grows with the command strength ||g||^2*Avar relative to the per-source noise; a threshold placed in the gap separates them. QED (asymptotic; finite-S concentration and the false-alarm rate are the subject of the synthetic study in Section 7 and the operational validation of Section 10).

COROLLARY (evasion cost, addresses A2). Driving lambda_ratio_coord down to the independent value 1/d requires ||g||^2*Avar -> 0, i.e. removing the shared command component. A botnet with no shared command component is not coordinated and loses the operational advantage of coordination. Evasion is therefore not free; it is paid in lost coordination. Section 5.6 (Theorem B5) sharpens this from an asymptotic statement into an exact spread-invariant identity.

5.2. Theorem B2 (Corroboration / False-Positive Bound)

STATEMENT. Suppose each of V vantages independently flags a given source as a candidate member with false-positive probability at most p (i.e., labels a benign source as a member with probability <= p). Require that a source be admitted to the identified set only if at least V vantages agree. Then:

(i) Under vantage independence, the host-level false-positive probability satisfies P_fp <= p^V.

(ii) Under partial correlation with pairwise correlation rho in [0,1], P_fp <= p^V + (1 - (1-rho)^(V-1)) * (p - p^V), which reduces to p^V at rho=0 and to p at rho=1.

PROOF. (i) is the product rule for independent events. (ii) interpolates: with probability (1-rho)^(V-1) the V-1 confirming judgements behave independently of the first (giving p^V), and otherwise they may collapse onto a single shared error (giving p); the convex combination yields the bound. QED.

REMARK. This is the closed-form expression of the qualitative requirement in RFC 6561 Section 4 ("a combination of multiple bot detection data points ... to avoid or minimize false-positive identification of hosts").

COROLLARY B2(iii) (independence is the precondition). The corroboration gain comes ENTIRELY from vantage independence. Under independence (rho=0), V=3 at p=0.05 already gives P_fp <= 1.25e-4, below a 1e-3 target. Under correlation, the bound of (ii) does NOT vanish with V: as V grows it converges UP to p, and for rho=0.1 it is floored at roughly 7e-3 (minimised near V=2). No number of CORRELATED vantages reaches a 1e-3 target. Operationally: the false-positive guarantee is only as strong as the path/observation diversity of the vantages. This is verified in validate_botnet_coherence.py check T-B2-3, and is exactly why a legitimate flash crowd -- whose per-vantage errors are correlated -- is the principal residual false positive (Section 7, Section 10).

5.3. Theorem B3 (No Unilateral Action / Remediation Soundness)

STATEMENT. Under this profile, no source’s host state or traffic is altered by MVPS. Every state-changing action is performed by an external control point governed by RFC 6561, DOTS, MUD, or BCP 38, each of which receives MVPS output as advisory input.

JUSTIFICATION. The detector's only output is a signed evidence object (Section 6.1). The hand-off interfaces of Section 6 are all request/advisory: an RFC 6561 process MAY notify; a DOTS server MAY mitigate; a MUD enforcement point MAY contain. Because MVPS holds no enforcement capability, no false positive at the detector can, by itself, quarantine a host -- it can only raise a corroborated request that the responsible, policy-bound control point evaluates. This is the property that makes the false-positive bound of Theorem B2 a SAFETY bound and not merely an accuracy figure.

5.4. Theorem B4 (Falsifiability / coherence-collapse axis)

MOTIVATION. Theorems B1-B2 rest on a premise that does NOT always hold: that a benign population's per-vantage errors are independent, so corroboration drives them away. The premise fails for a legitimate FLASH CROWD: a real, correlated event (a news spike, a software-update thundering herd) makes many sources move together in the human-driven features (arrival rate, destination), so the crowd looks coordinated across vantages and SURVIVES corroboration (observed directly in Section 7: flash-crowd corroborated false positive 0.115, exceeding the independence bound 0.035). This is the coherence environment "collapsing": cross-vantage agreement no longer certifies a botnet.

This is exactly the COHERENT_BUT_FALSE (CBF) failure mode of the MVPS AI-Coherence extension [I-D.melegassi-mvps-ai-coherence] (Sections 6-7 there): a consensus that is internally coherent yet wrong. In the botnet setting we call its dual COORDINATED-BUT-BENIGN (CBB). The AI-Coherence extension's answer to CBF is a falsifiability axis (its C_4): re-test the consensus on a dimension the failure mode cannot fake. We import that axis here.

STATEMENT. Partition the per-source feature space R^d into a HUMAN block H (features a legitimate crowd legitimately shares: arrival rate, destination entropy) and a MACHINE-REGULARITY block M (inter-arrival regularity, TTL stability, flag mix, payload-size dispersion -- features only a machine fleet shares). Write the shared command g = (g_H, g_M). Then the leading-eigenvalue ratio restricted to M:

- flash crowd (g_M = 0): the cross-source covariance on M is exactly sigma^2 I, so lambda_ratio|_M = 1/|M| -- it COLLAPSES to the independent floor, orthogonal to its full-space coordination (the exact analogue of C_4's orthogonality to C_1/C_2/C_3 for CBF, [I-D.melegassi-mvps-ai-coherence] Section 6.5);

- bot fleet (g_M != 0): lambda_ratio|_M >> 1/|M|.

ADMISSION RULE. Admit a source only if it is coordinated on the full space (B1) AND survives the falsifiability axis on M. This removes the flash-crowd residual that corroboration alone cannot.

PROOF. A flash crowd's shared component lies in H by hypothesis, so projecting out H leaves only independent per-source noise on M, whose covariance is sigma^2 I with all eigenvalues equal: lambda_ratio|_M = 1/|M| exactly. A bot fleet's command has g_M != 0, so the rank-1 term survives the projection and lambda_ratio|_M follows the B1 form on |M| coordinates. QED (closed form, validate_botnet_coherence.py checks T-B4-1..3).

EMPIRICAL CONFIRMATION (Section 7). On the machine-regularity subspace the flash crowd becomes indistinguishable from independent legitimate traffic (mean 0.299 vs 0.300) while the bot fleet stays high (0.746); adding the axis to the admission rule drives the flash-crowd corroborated false positive 0.115 -> 0.000 with botnet

detection held at 1.000.

COST AND LIMIT. The falsifiability axis costs the extra machine-regularity features per source; it does not defend against a future adversary that deliberately matches human-crowd statistics on M as well (the CBB analogue of a perturbation-stable hallucination, [I-D.melegassi-mvps-ai-coherence] Section 6.5 caveat). Such an adversary pays the full coordination-suppression cost of the B1 corollary on every monitored feature.

5.5. When vantages collapse: Byzantine-robust aggregation

B2-B4 assume the vantages themselves are honest-but-noisy. A compromised or hijacked vantage is a second way the coherence environment collapses: one Byzantine vantage can drag an arithmetic-mean centroid arbitrarily, forging or masking coordination. This profile inherits the cell-aware minimax Byzantine bound floor((k-1)/2) of [I-D.melegassi-mvps-ddos-resilience], and, where per-vantage distributions are aggregated, MUST use the geometric-median estimator C_2^gm of [I-D.melegassi-mvps-ai-coherence] (Section 11), whose breakdown point is 1/2 (versus 1/N for the mean), together with that document's SUSPECTED_BYZANTINE label for vantage attribution. MVPS still emits evidence only (Theorem B3); Byzantine robustness changes WHICH evidence is trustworthy, never whether MVPS acts.

5.6. Theorem B5 (No Free Decorrelation -- multi-vantage aggregation)

The B1 corollary states that evasion costs coordination, but only asymptotically. This theorem makes it exact and removes the last hope of an adversary: that spreading the coordination thinly enough across sources defeats detection.

STATEMENT. Model the botnet's coordinated effect as a coherent energy E that lives in the observable rowspace of the multi-vantage operator (the "observable = actuated" regime: a coordinated effect with no projection on the observation rowspace is, by definition, effect with no measurable consequence). An adversary that spreads E evenly across N sources/vantages drives the per-vantage signal to mu = sqrt(E/N) -> 0, but the multi-vantage coherent statistic is

$$T_{agg} = \sqrt{E}, \quad \text{INDEPENDENT of } N,$$

and contains NO computational-cost term. Consequently:

(i) no free stealth: $T_{agg}^2 = E$ exactly (what the dome sees equals the coordinated effect delivered);
(ii) the advantage $adv(N) = \sqrt{E} - \sqrt{E/N}$ is positive and strictly increasing in N -- the very spreading that blinds a single sensor WIDENS the multi-vantage lead;
(iii) silent-coordination cap: any botnet delivering coherent effect $E \geq \tau^2$ is detected with probability ~1 for ANY spread; the region {large coordinated effect, detector silent} is empty.

PROOF. Even spread gives per-source mean sqrt(E/N) over N terms; the coherent sum normalised by sqrt(N) has mean $\sqrt{N}\cdot\sqrt{E/N} = \sqrt{E}$ regardless of N. No step references compute, so a faster (including quantum) adversary cannot lower it: it is an information bound, not a computation. QED (closed form, validate_botnet_coherence.py checks T-B5-1..3).

ADVERSARIAL RED-TEAM. Letting an adversary actually play (scripts/simulate_botnet_redteam.py, Attack 1) confirms it: as the adversary spreads E across N = 1..128 vantages the single-vantage detection rate collapses 0.978 -> 0.0001 while the multi-vantage

aggregate stays at 0.977 with mean statistic fixed at sqrt(E) = 6.0. The spread that blinds one eye only widens the multi-vantage lead.

CONSEQUENCE FOR A2 (decorrelation evasion). The orthogonal decomposition stealth(a) + effect(a) = ||a||^2 (Pythagoras on the rowspace/nullspace split of the observation operator) is exact: every unit of coordination an adversary hides from the observable space is a unit of coordinated effect that no longer happens. "Hidden coordination is coordination that does nothing."

5.7. Theorem B6 (Non-Blinding of the corroboration set)

B2-B5 assume the vantages report honestly. The final evasion is to corrupt the vantages so the coordination becomes invisible -- to blind the detector. This theorem bounds it.

STATEMENT. Let V vantages observe a coordination subspace of effective dimension d_eff, with redundancy rho = V - d_eff. Then:

- (i) while at most k <= rho vantages are corrupted, the honest sub-operator keeps full column rank and EVERY coordination direction stays detectable -- silent blinding is impossible;
- (ii) blinding requires k > rho corruptions, which open a null space of dimension k - rho that the vantage-integrity monitor reports as a stealth-gap > 0 (the count of missing or invalid authenticated reports): "blind" implies "known-blind";
- (iii) the only UN-flagged corruption is forging authenticated vantage reports, gated by a post-quantum signature (ML-DSA, [FIPS204]) with forgery probability <= 2^-lambda.

Hence P(silent blinding) <= 0 (while rho >= 1) + 2^-lambda (PQC). Because the geometry term (i)-(ii) carries no computational variable, the bound holds against any future technology, including quantum; the composite inherits only the PQC exponent (ML-DSA-65 gives ~2^-112 over a generous 2^80 ten-year quantum query budget).

PROOF. Rank/SVD of the honest sub-operator (closed form, validate_botnet_coherence.py checks T-B6-1..3).

ADVERSARIAL RED-TEAM. Attack 2 of scripts/simulate_botnet_redteam.py corrupts k = 0..7 of V = 8 vantages (rho = 3): for k <= 3 the honest sub-operator keeps full column rank (stealth dimension 0, no blinding); for k > 3 a blinding null space of dimension k - 3 appears AND the stealth-gap reported by the integrity monitor equals it exactly (always flagged). Attack 3 confirms the Section 5.5 rule: a single Byzantine forgery of growing magnitude drags the arithmetic-mean centroid without bound (drift 1.4 -> 1397) while the geometric median stays bounded (~0.42).

DESIGN RULE (what this requires of a deployment). Ship V >= d_eff+1 DIVERSE vantages (diversity, not count, is what guarantees rank -- Section 2.6), authenticate every vantage report with a PQC signature ([FIPS204], replacing or wrapping the inherited Coherence-BFD AuthHMAC-SHA256 TLV 0xE9), and surface the stealth-gap as a first-class "known-blind" alarm. The promise is not "never blind" but "never SILENTLY blind".

6. From Evidence to Sanitization (the pipeline)

6.1. Evidence object (IoC, RFC 9424 framing)

For each admitted source, the broker produces an evidence object containing at least:

- source identifier (IP / prefix), and BCP 38 validation tag;
- lambda_ratio and D^2 time-series excerpts (the coordination signature, Section 4);
- lambda_ratio_machine and the falsifiability_pass flag (the machine-regularity-subspace falsifiability axis, Theorem B4);
- V (number of agreeing vantages) and the estimated P_fp (Theorem B2);
- observation window and a content hash;
- an HMAC/signature inherited from [I-D.melegassi-coherence-bfd] Section 12.

This object is a network-level IoC in the sense of RFC 9424 and is serialised into an IODEF [RFC7970] document (JSON binding [RFC8727]) for exchange.

The per-vantage observations the broker corroborates are ordinary flow records: IPFIX [RFC7011] export, whose Information Elements [RFC7012] (octet/packet counts, durations, ports, protocol) are the feature substrate of the coordination signature, including bidirectional flows exported as biflows [RFC5103]. No bespoke telemetry is required; a deployment corroborates over IPFIX collectors it already operates.

6.2. Hand-off to RFC 6561 remediation

For host-actionable sources (BCP 38 validated, not spoof-suspect), the evidence object is delivered to the operator's RFC 6561 process. That process -- NOT MVPS -- decides on notification, walled-garden placement, or other remediation, subject to RFC 6561's privacy and non-disruption requirements. The MVPS P_fp estimate SHOULD be carried so the RFC 6561 operator can apply its own confidence threshold.

6.3. Hand-off to DOTS / MUD / BCP 38

- Active flood, protected resource: broker MAY originate a DOTS [RFC9132] mitigation request scoped to the identified sources, and MAY pre-stage the coordination metrics (lambda_ratio, agreeing-vantages, p_fp_estimate) as DOTS telemetry [RFC9244] so the operator sees the corroborated evidence before any action.

- Constrained/IoT source with a MUD profile: deviation evidence is delivered to the MUD [RFC8520] enforcement point for containment.

- Spoof-suspect source (failed BCP 38 validation): NOT host-remediated; routed to traffic-level mitigation and reported to the upstream for ingress-filtering follow-up per BCP 84.

6.4. Canonical export: YANG and JSON

The evidence object of Section 6.1 has a canonical machine encoding, so it interoperates with model-driven and SIEM pipelines without a bespoke format. It follows the conventions of the MVPS telemetry export model [I-D.melegassi-opsawg-mvps-telemetry-export]:

- YANG module "catellix-mvps-botnet" (namespace urn:ietf:params:xml:ns:yang:mvps-botnet, prefix mvpsb) defines the read-only notification "mvps-botnet-evidence" carrying the Section 6.1 fields (source-address, bcp38-validated, lambda-ratio, d2, lambda-ratio-machine and falsifiability-pass (Theorem B4 axis), agreeing-vantages, p-fp-estimate, window, disposition -- including the coordinated-but-benign label -- content-hash, auth-hmac). The module is delivered via YANG-Push [RFC8641] over NETCONF/RESTCONF exactly as the MVPS telemetry export channel C. It is read-only: it carries

no command or actuation (Theorem B3).

- JSON Schema "mvps-botnet-evidence-v1" (2020-12) defines the equivalent JSON object. Its stable identifier is evidence_id = SHA-256(JCS(evidence \ {evidence_id})) per JCS [RFC8785], identical in spirit to the telemetry event_id, so producers are deterministic and consumers can deduplicate.

The JSON object is directly embeddable as a network-level Indicator of Compromise in an IODEF [RFC7970] document (JSON binding [RFC8727]); the lambda_ratio, agreeing-vantages, and p_fp_estimate travel as confidence metadata.

7. Results: detection on labelled ground truth (synthetic and real)

7.1. Synthetic labelled ground truth

PROVENANCE. The following are results of a LABELLED synthetic ground-truth experiment, not an operational botnet capture. You cannot honestly answer "did we find the botnet?" without ground truth, so every source is tagged at creation as one of three classes and the detector is scored against the labels it never sees. Reproducibility: scripts/simulate_botnet_coherence.py (seed 20260603, d = 6, V = 8 vantages, corroboration V_required = 3, 200 sources/population, 400 populations/class); receipt evidence/botnet_coherence_sim_receipt.json, body_sha256 460ccb48... The closed-form theorem checks are in scripts/validate_botnet_coherence.py (19/19 PASS, B1-B6), evidence/botnet_coherence_receipt.json, body_sha256 c1a2c31a...

Three classes: LEGIT (independent), FLASH CROWD (a legitimate event correlated in a few features -- the deliberate hard negative of Section 10), and BOTNET (one controller, shared command direction).

Coordination separation (Theorem B1), mean lambda_ratio (floor 1/d = 0.167):

Class	mean lambda_ratio
legit	0.213
flash crowd	0.290
botnet	0.662

Detection with V_required = 3 of 8 corroboration (threshold calibrated to a per-vantage p = 0.05):

```
botnet detection rate ........ 1.000  (400/400 admitted)
ROC AUC (botnet vs rest) ..... 1.000
legit false-positive rate .... 0.000
flash-crowd false-positive ... 0.115   (reported, not hidden)
overall false-positive ....... 0.0575
```

The honest, load-bearing finding (Theorem B2 / B2(iii)). The legit class's per-vantage errors are INDEPENDENT; corroboration collapses their population false positive to ~0, matching the p^V bound. The flash crowd's per-vantage errors are CORRELATED (the crowd shares a real signal), so they SURVIVE corroboration: measured flash-crowd population FP 0.115 EXCEEDS the independent binomial bound 0.035. This is B2(iii) observed directly: the corroboration guarantee requires vantage independence, and the flash crowd is the coherence-collapse case.

The falsifiability axis resolves it (Theorem B4). Re-testing the apparent coordination on the machine-regularity subspace M (the

AI-coherence axis of [I-D.melegassi-mvps-ai-coherence]) separates the crowd from the fleet:

Class	mean lambda_ratio|_M (floor 1/|M| = 0.250)	
legit	0.300	
flash crowd	0.299	(collapses to the legit floor)
botnet	0.746	(survives)

Admitting only sources that are coordinated (B1) AND survive the falsifiability axis (B4) gives:

```
flash-crowd false-positive ... 0.115 -> 0.000
botnet detection rate ........ 1.000  (unchanged)
```

So the residual that pure corroboration cannot remove is removed by the AI-coherence axis -- exactly the "coherence collapses, AI enters" structure of [I-D.melegassi-mvps-ai-coherence]. These numbers are synthetic and MUST be reproduced on operational data before any non-experimental claim; in particular an adversary that fakes human-crowd statistics on M as well is not defended (Section 5.4 cost-and-limit, Section 10). Sections 7.2 and 7.3 take the first step of that reproduction on real labelled traffic.

7.2. Real labelled botnet traffic (CTU-13) -- detection

PROVENANCE. The following are MEASURED on real labelled botnet traffic: the CTU-13 dataset of the Stratosphere IPS Laboratory, bidirectional flow records [RFC5103] labelled Botnet / Normal / Background. Five scenarios spanning three malware families were used: Neris (scenario 9), Rbot (scenarios 4, 11), and Virut (scenarios 5, 13). Reproducibility: scripts/collect_ctu13_botnet.py (download + label-preserving reduction; per-capture provenance and stream SHA in evidence/ctu13_raw/*_meta.json) and scripts/validate_ctu13_coordination.py; receipts evidence/ctu13_coordination_receipt_s*.json and the cross-family summary evidence/ctu13_coordination_combined_receipt.json (body_sha256 287f8bef...).

Two independent tests are run, deliberately, because the lab captures contain very few simultaneously-infected hosts (often one), which is too few to measure across-source coordination directly.

(a) Per-flow detectability (robust to host count). A held-out Fisher-LDA on the per-flow features (octets, packets, duration, source/total byte ratio, rate, protocol, destination port -- all IPFIX Information Elements [RFC7012]) separates Botnet from Normal flows with:

Family	Scenario	held-out AUC (Botnet vs Normal)
Neris	9	0.854
Rbot	4	0.966
Rbot	11	0.999
Virut	5	0.929
Virut	13	0.938

Mean AUC 0.937, minimum 0.854, replicated across three families. This is a real-data DETECTION result; it is not, by itself, a proof of the B1 coordination MECHANISM.

(b) Coordination signature (Theorem B1), where measurable. For the across-source leading-eigenvalue ratio lambda_ratio to be meaningful it MUST be compared to a same-size null, as lambda_ratio inflates when the source count is small. Drawing equally many random

non-botnet sources (2000 draws) gives a null whose 95th percentile is the bar to beat:

Scenario	#bot hosts	lambda_ratio	null p95	z vs null
9 Neris	10	0.813	0.711	+2.96
11 Rbot	3	1.000	0.967	+1.69
4,5,13	1	n/a	n/a	not testable

Where there are enough infected hosts to test (>= 3), the botnet ratio exceeds the same-size null -- i.e. it is NOT a small-sample artefact -- strongest in Neris at ~3 sigma. Where a scenario captured a single infected host, across-source coordination is not measurable and is reported as such rather than asserted.

HONESTY. This establishes (i) real-data detectability across families and (ii) a real, null-controlled B1 signature where the host count permits. It does NOT establish B2 multi-vantage corroboration on real data: CTU-13 is a single capture point (Section 10).

7.3. The multi-vantage advantage on measured real effect sizes

PROVENANCE. Measured on the same CTU-13 captures. Reproducibility: scripts/validate_mvps_advantage_ctu13_real.py; receipt evidence/mvps_advantage_ctu13_real_receipt.json (body_sha256 5f67a31d...).

Theorem B5 (No Free Decorrelation) states that the multi-vantage coherent statistic is spread-invariant with no compute term, so the detection advantage GROWS with aggregation. Until now this was shown only on the synthetic z-game. Here its INPUT -- the per-observation effect size delta -- is measured from real botnet flows (the held-out separation in normal-sigma units), and the prediction is checked empirically:

Family	delta (sigma)	single-flow AUC	coherent AUC (K=16)
Neris	1.43	0.854	0.9999
Rbot	19.96	0.999	1.000
Rbot	4.21	0.963	1.000
Virut	2.76	0.939	1.000
Virut	2.32	0.927	1.000

The empirical coherent statistic (aggregating K real flows) meets or exceeds single-flow detection for every family and approaches 1 as K grows -- the measured form of the B5 advantage. Instantiating the z-game with the measured delta, the number of coherent observations needed for >= 0.99 detection is 20 (Neris), 6-8 (Virut), and 1-3 (Rbot): even the weakest real family is decisively detected by a few coherent observations, while a single diluted vantage is not.

HONESTY. delta is empirical; the detection-rate model uses the Gaussian z-game convention (tau, sigma = 1) of the synthetic proof. This is a real-data INSTANTIATION of the advantage, not a B2 corroboration across independent real observers.

8. Security Considerations

This document introduces no new wire format or cryptographic primitive; transport security, authentication, replay protection, and control-plane isolation are inherited from [I-D.melegassi-coherence-bfd] and

[I-D.melegassi-mvps-ddos-resilience].

Misuse risk. A corroborated-evidence engine could be misused to justify wrongful blocking. Theorem B3 is the structural mitigation: MVPS cannot itself block, and the false-positive bound of Theorem B2 MUST be carried with every evidence object so the downstream control point can apply policy. Operators MUST NOT configure automatic host quarantine on single-vantage (V=1) evidence.

Adversarial decorrelation (A2) is bounded exactly by Theorem B5: spreading the coordination cannot hide its coherent energy from the multi-vantage aggregate, and the bound carries no compute term (AI/quantum cannot lower it). Vantage corruption / blinding (A4) is bounded by Theorem B6: silent blinding is impossible while the redundancy rho = V - d_eff >= 1 with diverse vantages, blinding above that is flagged ("known-blind"), and the only un-flagged corruption is PQC-gated forgery ([FIPS204], <= 2^-lambda). Adversarial control of a strict majority of vantages/cells remains out of scope and inherits the Byzantine bound floor((k-1)/2) of [I-D.melegassi-mvps-ddos-resilience]; Section 2.6 (RPKI/ROV, SAVI) is how rho is kept >= 1 in practice.

Evidence forgery is mitigated by the inherited HMAC and monotonic sequence numbers; an evidence object with a broken signature MUST be discarded and MUST NOT reach a remediation process.

9. Privacy Considerations

Identifying sources as botnet members is, by construction, the handling of data about individual endpoints, which may be personally identifiable. This profile therefore inherits the privacy requirements of RFC 6561 (Section 5 and its privacy discussion) and the framework of [RFC6973].

Specifically:

- Per-source evidence MUST be access-controlled and MUST NOT be published in raw form.
- When shared cross-organisation, evidence SHOULD carry the minimum necessary fields (source, confidence, window) and SHOULD follow IODEF [RFC7970] handling/marking.
- The coherence statistics MUST NOT carry user payload.
- Retention of per-source evidence SHOULD be bounded to the remediation window plus an audit period defined by operator policy.

10. Operational and Validation Considerations

This document is Experimental. Section 7.2/7.3 already take the first reproduction step on REAL labelled traffic (CTU-13, three families): real-data detectability (AUC 0.85-0.999) and a null-controlled B1 signature where the host count permits. The following remain REQUIRED before any progression or any non-experimental claim:

- Close Theorem B2 on real data: corroborate the SAME event across THREE OR MORE INDEPENDENT real vantages. CTU-13 is a single capture point, so it demonstrates detection and the B1 coordination signature but NOT multi-vantage corroboration. Suitable sources include a network telescope/darknet, multi-provider IPFIX [RFC7011] export, or an operator's own labelled incident observed from >= 3 collectors.
- Reproduce the B1 across-source signature on a capture with MANY simultaneously-infected hosts (CTU-13 lab scenarios have few), so the leading-eigenvalue test has high statistical power

across families, not only the ~3 sigma seen on Neris.
- Calibrate the per-vantage false-positive rate p and the vantage correlation rho on that data; Theorem B2 is only as good as the measured (p, rho).
- Confirm that the low-rank premise of Section 4 holds for real coordinated populations and does NOT spuriously hold for large flash-crowd legitimate events (the principal expected false-positive source). On synthetic ground truth the botnet is perfectly separated (ROC AUC 1.0, Section 7); the flash crowd's 0.115 corroborated false positive is closed by the falsifiability axis (Theorem B4) ON SYNTHETIC DATA, but the feature partition H/M and the machine-regularity thresholds MUST be calibrated and the closure REPRODUCED on operational traces. Whether a real adversary can match human-crowd statistics on the machine-regularity subspace M (defeating B4) is the open adversarial question.
- Where vantages may be compromised, deploy the geometric-median aggregation and SUSPECTED_BYZANTINE attribution of [I-D.melegassi-mvps-ai-coherence] (Section 5.5) and measure the realised breakdown fraction.
- Verify the BCP 38 tagging path end to end, since host-level remediation of a spoofed source is both useless and harmful.

Manageability: implementations SHOULD expose counters for sources_identified, mean_V_at_admission, estimated_P_fp, spoof_suspect_count, and evidence_objects_emitted.

11. IANA Considerations

All packet formats, TLVs, and code points are inherited from [I-D.melegassi-coherence-bfd] and [I-D.melegassi-mvps-ddos-resilience]; this document requests none.

This document requests, upon adoption, registration of the YANG module "catellix-mvps-botnet" (Section 6.4) in the "YANG Module Names" registry, with a namespace URI of the form urn:ietf:params:xml:ns:yang:mvps-botnet. Pending that assignment the module is non-normative, consistent with the export module of [I-D.melegassi-opsawg-mvps-telemetry-export].

12. References

12.1. Normative References

[I-D.melegassi-ippm-mvps-bundle]
Melegassi, L., "Multi-Vantage Path Synchrony Bundle Envelope and Vector Algebra", draft-melegassi-ippm-mvps-bundle-00, May 2026.

[I-D.melegassi-coherence-bfd]
Melegassi, L., "Coherence-BFD: Sub-Tick Coherence Detection over BFD Mechanisms", draft-melegassi-coherence-bfd-00, May 2026.

[I-D.melegassi-mvps-incremental-be]
Melegassi, L., "Incremental Bandwidth-Efficient Multi-Vantage Path Synchrony (BE-MVPS): Cell-Partitioned Coherence with epsilon-Gated Sherman-Morrison Updates", draft-melegassi-mvps-incremental-be-00, May 2026.

[I-D.melegassi-mvps-ddos-resilience]
Melegassi, L., "Volume-Independent DDoS Detection via Coherence-BFD: The MVPS DDoS Resilience Profile", draft-melegassi-mvps-ddos-resilience-00, May 2026.

[I-D.melegassi-mvps-ai-coherence]

Melegassi, L., "MVPS AI-Coherence Extension: Semantic, Byzantine, and Infrastructure-Cognitive Coherence for AI-Serving Network Deployments", draft-melegassi-mvps-ai-coherence-01, May 2026.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2827] Ferguson, P. and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing", BCP 38, RFC 2827, May 2000.

[RFC3704] Baker, F. and P. Savola, "Ingress Filtering for Multihomed Networks", BCP 84, RFC 3704, March 2004.

[RFC6561] Livingood, J., Mody, N., and M. O'Reirdan, "Recommendations for the Remediation of Bots in ISP Networks", RFC 6561, March 2012.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, May 2017.

12.2. Informative References

[RFC4732] Handley, M., Ed., Rescorla, E., Ed., and IAB, "Internet Denial-of-Service Considerations", RFC 4732, November 2006.

[RFC5103] Trammell, B. and E. Boschi, "Bidirectional Flow Export Using IP Flow Information Export (IPFIX)", RFC 5103, January 2008.

[RFC6480] Lepinski, M. and S. Kent, "An Infrastructure to Support Secure Internet Routing", RFC 6480, February 2012.

[RFC6545] Moriarty, K., "Real-time Inter-network Defense (RID)", RFC 6545, April 2012.

[RFC6811] Mohapatra, P., Scudder, J., Ward, D., Bush, R., and R. Austein, "BGP Prefix Origin Validation", RFC 6811, January 2013.

[RFC7011] Claise, B., Ed., Trammell, B., Ed., and P. Aitken, "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information", STD 77, RFC 7011, September 2013.

[RFC7012] Claise, B., Ed. and B. Trammell, Ed., "Information Model for IP Flow Information Export (IPFIX)", RFC 7012, September 2013.

[RFC7039] Wu, J., Bi, J., Bagnulo, M., Baker, F., and C. Vogt, Ed., "Source Address Validation Improvement (SAVI) Framework", RFC 7039, October 2013.

[RFC8210] Bush, R. and R. Austein, "The Resource Public Key Infrastructure (RPKI) to Router Protocol, Version 1", RFC 8210, September 2017.

[FIPS204] National Institute of Standards and Technology, "Module-Lattice-Based Digital Signature Standard (ML-DSA)", FIPS 204, August 2024.

[RFC6973] Cooper, A., Tschofenig, H., Aboba, B., Peterson, J., Morris, J., Hansen, M., and R. Smith, "Privacy Considerations for Internet Protocols", RFC 6973,

July 2013.

[RFC7970] Danyliw, R., "The Incident Object Description Exchange Format Version 2", RFC 7970, November 2016.

[RFC8520] Lear, E., Droms, R., and D. Romascanu, "Manufacturer Usage Description Specification", RFC 8520, March 2019.

[RFC8727] Takahashi, T., Suzuki, M., and R. Danyliw, "JSON Binding of the Incident Object Description Exchange Format", RFC 8727, August 2020.

[RFC8783] Boucadair, M., Ed. and T. Reddy.K, Ed., "Distributed Denial-of-Service Open Threat Signaling (DOTS) Data Channel Specification", RFC 8783, May 2020.

[RFC8811] Mortensen, A., Reddy.K, T., and R. Moskowitz, "DDoS Open Threat Signaling (DOTS) Architecture", RFC 8811, August 2020.

[RFC9132] Boucadair, M., Ed., Shallow, J., and T. Reddy.K, "Distributed Denial-of-Service Open Threat Signaling (DOTS) Signal Channel Specification", RFC 9132, September 2021.

[RFC9244] Boucadair, M., Ed., Reddy.K, T., Ed., Doron, E., Chen, M., and J. Shallow, "Distributed Denial-of-Service Open Threat Signaling (DOTS) Telemetry", RFC 9244, June 2022.

[RFC8641] Clemm, A. and E. Voit, "Subscription to YANG Notifications for Datastore Updates", RFC 8641, September 2019.

[RFC8785] Rundgren, A., Jordan, B., and S. Erdtman, "JSON Canonicalization Scheme (JCS)", RFC 8785, June 2020.

[RFC9424] Paine, K., Whitehouse, O., Sellwood, J., and A. Shaw, "Indicators of Compromise (IoCs) and Their Role in Attack Defence", RFC 9424, August 2023.

[I-D.melegassi-opsawg-mvps-telemetry-export] Melegassi, L., "Exporting MVPS Coherence Events over Standard Telemetry Channels (syslog, IPFIX, YANG-Push)", draft-melegassi-opsawg-mvps-telemetry-export-00, May 2026.

[I-D.melegassi-opsawg-mvps-yang-model] Melegassi, L., "A YANG Data Model for Multi-Vantage Path Snapshots (MVPS)", draft-melegassi-opsawg-mvps-yang-model-00, May 2026.

[I-D.melegassi-mvps-perfsec-coupling] Melegassi, L., "MVPS Performance-Security Coupling Profile", draft-melegassi-mvps-perfsec-coupling-00, May 2026.

[I-D.melegassi-santos-ippm-mvps-cwt] Melegassi, L. and Santos, "MVPS Trust Profile: Coherent-Witness Trust (CWT)", draft-melegassi-santos-ippm-mvps-cwt-00, May 2026.

Appendix A. Reproducibility (validators, simulations, receipts)

Every claim in this document is either an algebraic identity checked by a validator, or a labelled simulation with a signed receipt.

```
   Theorem checks (closed form), B1-B6:
     scripts/validate_botnet_coherence.py    19/19 PASS, exit 0
     evidence/botnet_coherence_receipt.json   body_sha256 c1a2c31a...

   Labelled ground-truth detection experiment (incl. the B4
   falsifiability-axis resolution of the flash-crowd residual):
     scripts/simulate_botnet_coherence.py
     evidence/botnet_coherence_sim_receipt.json
                                              body_sha256 460ccb48...
     docs/SIM_BOTNET_RESULTS.txt

   Adversarial red-team, Monte-Carlo companion to B5/B6
   (spreading, blinding, Byzantine centroid -- all defended):
     scripts/simulate_botnet_redteam.py
     evidence/botnet_redteam_receipt.json
                                              body_sha256 d9a59fd8...
     docs/SIM_BOTNET_REDTEAM_RESULTS.txt

   Canonical export:
     schema/catellix-mvps-botnet.yang         (notification model)
     schema/mvps-botnet-evidence.schema.json  (JSON Schema 2020-12)
     evidence/botnet_evidence_example.json    (worked instance)

   Real labelled botnet traffic (CTU-13, Stratosphere IPS Lab;
   Section 7.2) -- collection, detection, and B1 with same-size null
   (scenarios 9 Neris; 4,11 Rbot; 5,13 Virut):
     scripts/collect_ctu13_botnet.py
     evidence/ctu13_raw/ctu13_s*_meta.json  (source URL + stream SHA)
     scripts/validate_ctu13_coordination.py
     evidence/ctu13_coordination_receipt_s9.json   a046e86a...
     evidence/ctu13_coordination_receipt_s11.json  e26fb4f3...
     evidence/ctu13_coordination_receipt_s4.json   57b9b36c...
     evidence/ctu13_coordination_receipt_s5.json   94bc9038...
     evidence/ctu13_coordination_receipt_s13.json  49a5fe3d...
     evidence/ctu13_coordination_combined_receipt.json
                                           body_sha256 287f8bef...

   Multi-vantage advantage (Theorem B5) instantiated on measured
   real effect sizes (Section 7.3):
     scripts/validate_mvps_advantage_ctu13_real.py
     evidence/mvps_advantage_ctu13_real_receipt.json
                                           body_sha256 5f67a31d...

   Honest negative (free public threat feeds observe DISJOINT
   populations and so cannot, alone, corroborate B2 on real data --
   the basis for the Section 10 requirement):
     scripts/collect_threat_feeds.py
     scripts/validate_real_botnet_coherence.py
     evidence/real_botnet_coherence_receipt.json
                                           body_sha256 455bc967...

The CTU-13 dataset is the labelled botnet corpus of Garcia et al.,
"An empirical comparison of botnet detection methods", Computers &
Security, 2014, distributed by the Stratosphere IPS Laboratory at
https://www.stratosphereips.org/datasets-ctu13.  The collector
records the exact per-capture source URL and a streaming SHA-256 in
evidence/ctu13_raw/*_meta.json so the reduction is auditable; raw
per-source evidence is kept private per Section 9 and is NOT part of
any public page.

The body_sha256 of each receipt is computed over the JCS [RFC8785]
serialization of the receipt body BEFORE any environment-specific
field is attached, so it reproduces bit-for-bit on any machine.
```

Appendix B. Implementation and Deployment Guidance

This appendix is informative. It describes a minimal, standards-based way to implement the profile; it adds no normative requirement beyond the body of the document.

B.1. Components

- Vantages (>= V, diverse): existing flow exporters. Each emits IPFIX [RFC7011] records (bidirectional biflows [RFC5103] are preferred) and signs its telemetry with a post-quantum identity (ML-DSA, [FIPS204]) so the non-blinding gate of Theorem B6 holds.
- Broker: collects per-vantage records, computes the coherence and coordination statistics, applies corroboration, and emits evidence objects (Section 6.1). The broker NEVER actuates (Theorem B3).
- Control points: the operator's existing RFC 6561 process, DOTS [RFC9132]/[RFC9244] server, MUD [RFC8520] enforcement point, and BCP 38 ingress filters. These -- not the broker -- remediate.

B.2. Per-window computation (the broker)

For each observation window of width (M-1)*T_tick:

1. Ingest IPFIX records per vantage; the features are the Information Elements [RFC7012] of Section 7.2 (octets, packets, duration, source/total byte ratio, rate, protocol, destination port).
2. Per candidate source, form its feature vector and score it on each vantage; a per-vantage flag uses the calibrated per-vantage false-positive rate p (Section 7.1, p = 0.05 in reference run).
3. Coordination test (Theorem B1): assemble the across-source matrix and compute the leading-eigenvalue ratio lambda_ratio. Compare it to a SAME-SIZE null (equally many non-candidate sources, >= 1000 draws); admit the coordination signal only if it exceeds the null 95th percentile. This null control is mandatory because lambda_ratio inflates for small source counts (Section 7.2).
4. Corroboration (Theorem B2): admit a source only if V_required of V independent vantages agree (V_required = 3 in the reference run). Carry the estimated P_fp and the measured vantage correlation rho.
5. Falsifiability re-test (Theorem B4): for sources that look coordinated, re-run step 3 on the machine-regularity subspace M; a flash crowd collapses to the independent floor and is dropped, a bot fleet survives.
6. Integrity (Theorem B6): compute the redundancy rho = V - d_eff; if rho < 1, raise a "known-blind" alarm rather than emitting silent results; verify each vantage's [FIPS204] signature and discard forged or unsigned reports.

B.3. Calibration

The bounds are only as good as their measured inputs. Before production, calibrate on local data: the per-vantage p, the vantage correlation rho (Theorem B2 degrades from p^V toward the stated mixture bound as rho rises), the H/M feature partition and the machine-regularity thresholds of Theorem B4, and the detection threshold tau. Section 7.3 shows how the per-flow effect size delta maps to the number of coherent observations needed for a target detection probability.

B.4. Export and hand-off

Emit each admitted source as the Section 6.1 evidence object: the YANG notification "mvps-botnet-evidence" via YANG-Push [RFC8641], or the equivalent JSON (stable evidence_id = SHA-256(JCS [RFC8785])), embeddable as an IoC [RFC9424] in an IODEF [RFC7970] document (JSON binding [RFC8727]). Route per Section 6.2/6.3: host-actionable and

BCP 38-validated sources to the RFC 6561 process; active floods to a DOTS [RFC9132] request with coordination metrics pre-staged as DOTS telemetry [RFC9244]; constrained devices to their MUD [RFC8520] enforcement point; spoof-suspect sources to traffic-level mitigation and BCP 84 follow-up, NEVER to host remediation.

B.5. Manageability

Expose the counters of Section 10 (sources_identified, mean_V_at_admission, estimated_P_fp, spoof_suspect_count, evidence_objects_emitted) plus the redundancy rho and the known-blind alarm state, so operators can see when the corroboration guarantee is in force and when it is not.

Acknowledgements

This profile was prompted by the observation that the MVPS DDoS profile attributes an attack to a region but stops short of identifying participating sources, and by the recognition that RFC 6561 already names multi-point corroboration -- with false-positive avoidance -- as the hard part of bot remediation. The author thanks the IETF OPSEC, DOTS, and MILE communities for the standards this document is careful to build on rather than duplicate.

Author’s Address

Leonardo Melegassi
Catellix
Andradina, SP
Brazil

Email: melegassi@catellix.com
URI: https://catellix.com/