Attention Protocol

A Refundable-Deposit Standard for Pricing Human and Machine Attention

White Paper — Draft v0.1

July 2026

“Unknown senders should have to demonstrate that their message is worth the recipient’s attention.”

Thomas · attn.email · Draft for discussion — not an offer of securities or investment advice

Abstract

Digital communication has a structural flaw: the cost of demanding someone's attention is borne almost entirely by the person whose attention is demanded. Email, the clearest case, allows a sender to reach any recipient at effectively zero marginal cost, transferring the full cost of evaluation to the receiver. Every filtering technology built in response — spam detection, priority inboxes, AI triage — treats the symptom while leaving the economics intact, and the arrival of capable AI agents is about to make the imbalance categorically worse: personalized outreach at machine scale, screened by machines, paid for by no one.

This paper proposes a simple economic correction: the ***refundable attention deposit****. A recipient publishes, in machine-readable form, the price and rules under which unknown parties may claim their priority attention. A first-time sender — human or AI agent — attaches a small refundable deposit to their message. If the message is worth the recipient's time, the deposit is returned; if not, the recipient keeps it. Messages never reviewed are refunded automatically. We specify the* ***Attention Manifest****, an open discovery and payment schema that lets any inbox, service, or agent advertise its attention terms, settled by card for human senders and by stablecoin over the x402 standard for machine senders. We describe the incentive properties of the mechanism, a reference architecture, the trust-and-safety constraints that make it socially acceptable, and a deployment path that begins with a hosted service and ends with an open protocol.*

1 Introduction: The Economics of Free Attention

Attention is the scarcest resource in knowledge work, and it is the only scarce resource that strangers may consume without payment or permission. Sending an email costs fractions of a cent; reading, evaluating, and dismissing it costs the recipient seconds to minutes of finite time. Multiplied across a professional's inbound volume, this asymmetry produces the familiar condition of the modern inbox: important first-contact messages — the founder reaching an investor, the candidate reaching a hiring manager, the journalist reaching a source — drown in cold sales, recruiting spam, and automated pitches that were rational for their senders to send precisely because sending is free.

The standard response has been technological: Bayesian filters, sender reputation, machine-learned priority inboxes, and most recently large-language-model triage. These tools reduce noise but cannot eliminate it, because they do not alter the sender's incentive. A campaign with a 0.1% response rate remains profitable at zero marginal cost. Worse, filtering invites an arms race: as screening models improve, generation models improve against them, and both sides now use the same underlying technology. There is no equilibrium in which software alone wins, because every software defense can be probed and bypassed for free.

An economic constraint cannot. If reaching an unknown recipient requires posting even one dollar of refundable value, mass low-relevance outreach becomes expensive in exact proportion to its volume, while a single legitimate message costs its sender nothing but float. The idea is old (Section 2). What is new is the

environment: for the first time, machine senders have native payment rails. The x402 standard has made stablecoin payment a first-class HTTP operation, processing over 165 million agent transactions in its first months under Linux Foundation governance, with Stripe and Cloudflare support. Agents can now pay. Nothing yet tells them what a human's attention costs, or under what rules it may be claimed. Defining that missing recipient-side layer is the purpose of this protocol.

2 Prior Art and Why It Failed

The intellectual lineage of attention pricing is substantial, and honesty about its failures is a design requirement, not a rhetorical concession.

2.1 Proof-of-work postage

Back's Hashcash (1997) and Microsoft's Penny Black project proposed making senders pay in computation rather than money. The insight — impose asymmetric cost on the sender — was correct. The implementation failed on two fronts: computation is not transferable (the recipient gains nothing), and hardware asymmetry meant spammers with botnets paid less per message than legitimate users on laptops. A cost that does not compensate the injured party and does not scale with the sender's resources prices attention badly.

2.2 Attention bonds

Loder, Van Alstyne and Wash (2006) formalized the *attention bond mechanism*: senders post a bond the recipient may seize if the message proves unwanted, and showed it can outperform both filtering and flat postage in welfare terms, because it uses the recipient's own judgment as the classifier. The mechanism was economically sound and commercially stranded: in 2006 there was no practical way to escrow, refund, and settle millions of sub-dollar bonds, no identity layer for senders, and no deployment path that did not require replacing the recipient's email provider.

2.3 Paid-introduction marketplaces

21.co, relaunched as Earn.com, built a marketplace where senders paid listed recipients to answer messages, and was acquired by Coinbase in 2018; the inbox product was subsequently wound down. Three structural lessons follow. First, a *marketplace* of recipients has a two-sided cold-start problem — the product was only useful if the person you wanted was listed. Second, non-refundable payment for replies framed the product as a toll on access, socially awkward for both parties. Third, sender demand was capped by human patience: only humans wrote messages, and humans dislike paying strangers to read them.

2.4 Design consequences

The protocol described here differs from each predecessor on the axis that killed it. Against proof-of-work: value transfers to the recipient and scales in currency, not hardware. Against attention bonds' deployment problem: settlement rails (Stripe Connect for fiat, x402/USDC for machines) now exist as commodities, and the alias architecture (Section 5.4) requires no change of email provider. Against the marketplace model:

each recipient's protected address is useful to them alone from day one — there is no directory to populate — and refund-by-default reframes payment as a bond behind the claim "this message is worth your time," not a fee for audience. And against the demand cap: machine senders, absent in every prior attempt, are the marginal buyer this time.

3 The Mechanism: Refundable Attention Deposits

3.1 Definition

A recipient *R* publishes attention terms: a deposit amount *d* (typically $1–$25), an expiration window *T* (default 14 days), and disclosed settlement rules. When an unknown sender *S* — one with no prior correspondence, allowlist entry, or completed deposit — contacts *R*, the message is held and *S* is challenged. If *S* pays *d*, the message enters *R*'s priority queue. *R* then takes exactly one of four actions: **refund** (message was legitimate; *S* is made whole), **retain** (message consumed attention without value; *R* keeps *d* minus platform fee), **block** (with either settlement), or **nothing** — in which case the deposit refunds automatically at *T*.

3.2 What payment buys

Payment guarantees transparent priority treatment: placement in the designated queue, notification under the recipient's settings, and observable settlement status. It never guarantees a reply, an agreement, a meeting, or an outcome. This distinction is load-bearing. A system that sells replies is a marketplace for access, with the ethical and regulatory exposure that implies; a system that sells honest prioritization under disclosed rules is infrastructure.

3.3 Incentive analysis

- **Bulk sender.** A campaign of *n* messages now carries capital-at-risk *n·d*, with recovery only when recipients individually judge messages valuable. At *d* = $2 and a realistic retention rate for unsolicited outreach, expected cost per thousand messages moves from ~$0 to thousands of dollars — spam becomes negative-expected-value at exactly the volumes that make it spam.
- **Legitimate stranger.** Expects a refund, so expected cost approaches zero plus float and any disclosed processing fee. The deposit functions as a credible signal — costly to fake at scale, cheap when the claim is true — which is the classic structure of a screening mechanism.
- **Recipient.** The dangerous incentive is harvesting: collecting deposits without reviewing messages. The auto-refund rule removes it — an unreviewed message always refunds, so a recipient earns only by doing the work of review. Retention additionally requires an affirmative action after genuine review, never a tracking pixel. Review-rate monitoring and payout holds (Section 7) enforce the constraint against adversarial recipients.
- **Platform.** Fees are charged only on retained deposits, so platform revenue scales with attention actually protected rather than with gross message flow — the platform has no incentive to maximize challenge volume.

3.4 Refundability as the social contract

The refund default is not a pricing detail; it is what makes the mechanism acceptable. Deposits refunded on reply, refunded on expiry, and retained only after review keep the system legible as an anti-spam bond rather than a celebrity tollbooth. The health metric of the network is therefore inverted from a naive marketplace: a *high* refund rate among paying senders is success, because it means the mechanism is admitting legitimate strangers and deterring the rest before they pay.

4 The Attention Manifest (Specification Draft)

The Attention Manifest is a small, versioned, permissively licensed convention by which any recipient — person, service, or agent — advertises attention terms in machine-readable form. It deliberately specifies only discovery, terms, and payment entry points; settlement logic, reputation, and fraud control are implementation concerns.

4.1 Discovery

- Domains and services: `https://{domain}/.well-known/attention.json`, optionally mirrored in a DNS TXT record at `_attention.{domain}`.
- Individuals without domains: a conforming provider hosts the manifest at a stable handle URL (e.g. `https://attn.email/u/{handle}/manifest`).

4.2 Schema (illustrative)

```
{ "version": "0.1",
  "recipient": "thomas@attn.email",
  "verified": true,
  "pricing": [
    {"category": "default", "amount": "2.00", "currency": "USD"},
    {"category": "agent",   "amount": "1.00", "currency": "USD"}],
  "refund_policy": {
    "unopened_auto_refund_days": 14,
    "reply_auto_refund": true,
    "review_required_to_retain": true},
  "guarantees": ["priority_placement", "settlement_status"],
  "non_guarantees": ["reply", "outcome", "response_time"],
  "payment": {
    "hosted_checkout": "https://attn.email/pay/{token}",
    "x402": {"endpoint": "https://api.attn.email/v1/contact/{handle}",
             "networks": ["solana", "base"], "asset": "USDC"}} }
```

Figure 1 — Attention Manifest v0.1, illustrative instance.

4.3 The challenge as HTTP 402

For machine senders the challenge is not an email but a status code. An agent POSTs a message to the recipient's contact endpoint and receives `402 Payment Required` carrying x402 payment requirements derived from the manifest. The agent attaches a USDC payment header, retries, and receives `202 Accepted`

with a challenge identifier and status URL; settlement outcomes (refund, retention, expiry) are delivered by webhook or polled. The entire machine integration is standard HTTP plus standard x402 — the protocol introduces no proprietary payment semantics. For human senders, the same challenge renders as an automated email with a hosted card checkout; the manifest is the single source of truth behind both surfaces.

5 Reference Architecture

5.1 Three layers

The system is intentionally stratified so that openness and revenue do not compete. **Layer 1** is the open specification above — free to implement, designed to be embedded in agent frameworks and tool registries. **Layer 2** is a hosted implementation: alias infrastructure, policy engine, escrow and settlement, refunds, fraud control, reputation, dashboards, and payouts — everything operationally hard, and the commercial product. **Layer 3**, deferred until Layers 1–2 achieve adoption, is protocol-native settlement: program-controlled escrow on existing rails, portable sender reputation, and independent conforming providers. No protocol token exists at any layer; the asset being priced is attention, and speculative instruments would damage both credibility and incentives.

5.2 Settlement rails

Fiat deposits settle through a marketplace payment provider (Stripe Connect): funds are charged at checkout, held in regulated payment infrastructure, and released by the settlement decision; recipients never directly custody sender funds, and the platform avoids money-transmitter exposure by construction. Machine deposits settle in USDC via existing x402 facilitators on Solana and Base. A unified ledger reconciles both rails; every financial action is idempotent and audit-logged.

5.3 Known-sender determination

A sender is known — and passes free — given any of: prior outbound correspondence from the recipient, presence in the recipient's imported contacts or allowlist, an allowlisted domain, membership in an accepted thread, or a previously completed valid deposit. False positives are treated as a first-class failure mode: challenging an existing contact is grounds for mandatory refund and is tracked as a headline quality metric (target < 2%).

5.4 The alias deployment path

The initial deployment attaches to no email provider's internals. Each recipient claims a protected address (e.g. `{handle}@attn.email`) and publishes it as their public front door; known senders forward through untouched, unknown senders are challenged, and paid messages forward into the recipient's real inbox with the deposit marked. This is provider-agnostic, requires no privileged mailbox scopes, and — usefully — makes the protected address and the manifest handle the same identity for human and machine senders alike. Deep provider integrations are enhancements, not prerequisites.

6 Machine Senders and the Agent Economy

Agent-to-human contact is the protocol's expansion joint. Autonomous systems increasingly need a human: to review an exception, authorize a decision, answer an expert question, or grant an interview. Today those requests arrive as unauthenticated email indistinguishable from spam, or do not arrive at all. Under this protocol an agent resolves the recipient's manifest, pays the posted deposit, and receives an accountable, refundable, statused channel to a human — while the human receives machine outreach that is rate-limited by money rather than by patience.

Three properties matter for the machine case. *Composability:* the flow is plain HTTP plus x402, so a client is trivially wrapped as a tool (e.g. an MCP server exposing `contact_human`) inside any agent framework. *Accountability:* agent requests carry developer identity today and, as the Agent Payments Protocol (AP2) matures, cryptographically signed mandates demonstrating that a human principal authorized the outreach — the natural spam filter of the agent era. *Symmetry:* nothing in the manifest restricts recipients to humans; a service or an agent may publish its own attention terms, making the same mechanism serve machine-to-machine prioritization. The protocol takes no position on how large agent-to-human volume becomes; it is designed so that the marginal cost of supporting the machine case, given the human case, is approximately zero.

7 Trust, Safety, and Prohibited Uses

A market for attention creates abuse surfaces on both sides, and the protocol's legitimacy depends on constraining its own participants.

- **Recipient abuse** — harvesting deposits, inflated pricing, misrepresented guarantees — is constrained by automatic refund of unreviewed messages, review-rate and deposit-to-open monitoring, payout holds and reserves, price caps pending verification, and suspension with mandatory refunds.
- **Sender abuse** — card fraud, harassment financed by deposits, block evasion through repayment, phishing and malicious attachments — is constrained by payment-risk screening, identity fingerprinting, block enforcement scoped to identity and payment instrument, attachment scanning, and rate limits. Payment must never launder a threat into a priority queue; content that violates law or platform policy is removed regardless of deposit, with the deposit's disposition following the abuse policy rather than the recipient's discretion.
- **Structural misuse** — the mechanism must not become a bribery channel or a means of monetizing access one has no right to sell. Categories requiring restriction or prohibition include public officials monetizing constituent access, fiduciaries and regulated professionals charging where ethics rules forbid it, employees monetizing employer-owned inboxes without authorization, and any participation by minors.

Messages of certain classes — legal notices, emergency and safety communications — must have a payment-free path, and the specification reserves a manifest field for declaring it. A protocol that priced a subpoena or a safety warning would deserve the criticism it received.

8 Network Economics

The hosted implementation charges a commission (on the order of 10%) on deposits recipients retain, a metered fee on machine-contact volume beyond a free tier, and subscriptions for advanced recipient features. Refunded deposits return to senders in full, with any processing cost either absorbed or separately and visibly disclosed — a deposit advertised as fully refundable must be exactly that. Because commission attaches only to retained deposits, and retained deposits are (by design) the minority outcome, per-recipient deposit revenue is modest; the compounding assets are the recipient network, the settlement history, and sender reputation — data that improves pricing, fraud control, and admission decisions for every participant and that no fork of the schema inherits. The specification remaining open is therefore not altruism but strategy: adoption of the schema by agent frameworks routes demand toward conforming providers, of which the reference implementation intends to be the best rather than the only.

9 Limitations and Open Questions

- **Adoption asymmetry.** The mechanism requires recipients to publish a protected address and senders to tolerate a challenge. Both are behavior changes; the concierge and beta phases exist to measure them rather than assume them, with pre-registered thresholds for sender conversion and recipient review rates.
- **Agent demand is young.** A substantial share of early x402 volume appears to be testing rather than commerce. The protocol treats machine senders as the expansion case, not the base case; the human funnel must justify the product on its own.
- **Platform response.** A major email or model provider could ship a captive variant. The protocol's defense is neutrality and speed — a provider-agnostic standard published first — and the observation that captive variants validate the category while remaining unattractive to every competitor of their owner.
- **Distributional concerns.** Pricing attention risks advantaging the wealthy sender. Mitigations — low default deposits, refund-by-default, charitable routing of retained deposits, and free passage for known senders — are design commitments, but the tension is real and is better acknowledged than argued away.
- **Jurisdictional questions.** Marketplace payment structuring, stablecoin settlement classification, taxation of retained deposits, and communication-law compliance vary by jurisdiction and require counsel before general availability.

10 Conclusion

Two facts define the moment. Attention is the binding constraint on knowledge work, and for the first time every class of sender — including software — can pay. The refundable attention deposit converts these facts into a small, comprehensible market: strangers post a bond behind the claim that their message is worth your time, honest claims cost nothing, dishonest ones fund the person they imposed on, and machines participate

through the same open standard as people. The Attention Manifest is deliberately modest — a discovery document, a schema, and a status code — because the history of this idea shows that the economics were never the hard part; deployment was. The rails now exist. What remains is to publish the standard, ship a trustworthy implementation, and let recipients decide what their attention is worth.

References

Back, A. (2002). *Hashcash — A Denial of Service Counter-Measure.* Technical report (scheme first announced 1997).

Loder, T., Van Alstyne, M., & Wash, R. (2006). An Economic Response to Unsolicited Communication. *Advances in Economic Analysis & Policy*, 6(1).

Coinbase Developer Platform (2025). *x402: An Internet-Native Payment Protocol.* Whitepaper; protocol subsequently stewarded by the x402 Foundation under the Linux Foundation (2026).

Google et al. (2025). *Agent Payments Protocol (AP2).* Specification announcement with 60+ collaborating payment and platform organizations.

Dwork, C., & Naor, M. (1992). Pricing via Processing or Combatting Junk Mail. *CRYPTO '92.*

This draft is published for discussion. It describes an engineering and product proposal; it is not legal, tax, or investment advice, and nothing herein constitutes an offer of securities.