Network Working Group
Internet-Draft
Intended status: Informational
Expires: 19 July 2026

E. Fjeldstrom
Independent

15 January 2026

A Systemic Meditation on Internet Connectivity Equilibrium
draft-fjeldstrom-meditation-on-connectivity-01

Abstract

This document presents a systemic meditation on how the Internet arrived at its present connectivity equilibrium. The analysis proceeds by retrospective reconstruction: examining observable adaptations, constraints, and deferred decisions across multiple layers of the stack, rather than by benchmarking, simulation, or protocol comparison.

The term "meditation" is used deliberately to indicate a method grounded in historical observation, accumulated operational experience, and the interpretation of persistent compensatory mechanisms as empirical evidence of structural conditions. The document does not assign fault, advocate specific remedies, or propose new protocol mechanisms. Instead, it seeks to explain how a sequence of locally rational responses to real pressures interacted over time to produce a stable, but heavily mediated, connectivity equilibrium at Internet scale.

Table of Contents

1. Purpose and Scope 3
1.1. Central and Subsidiary Theses 4
1.1.1. Central Thesis 4
1.1.2. Subsidiary Thesis 4
1.1.3. Scope Clarification 5
1.1.4. Clarifying Observation (Ambient Endpoints) 5
1.2. Ambient Endpoints and Their Progressive Withdrawal (By Layer) 5
2. Baseline Assumptions and Early Operating Conditions 7
2.1. Early RFC Evidence, Grouped by Theme 7
2.1.1. Mediation, Local Control, and Administrative Boundaries 7
2.1.2. Identity, Accountability, and the Meaning of "Free" 8
2.1.3. Fragmentation, Heterogeneous Environments, and Why "Normal" Features Were Deferred 8
2.1.4. Physical Reality, Delay, and Layer Blurring 9
2.1.5. Cost, Noise, Control-Plane Externalities, and the Turn Toward Managed High-Bandwidth Networks 9
3. Emergence of Existential Stressors 11
3.1. Fragmentation and Administrative Plurality 11
3.2. Physical Distance, Delay, and Interaction Breakdown . . . 11
3.3. Cost and Host Resource Exhaustion 11
3.4. Background Traffic and Unattributed Load 11
3.5. Unconditional Acceptance and Denial of Service 12
3.6. Routing Scale, Control-Plane Costs, and Exit-Gateway Geometry 12
3.7. Security Normalization: Routing Withdrawal, Filtering, and Firewalls 12
3.8. Evolving Internet Membership: From IP Reachability to Application-Level Participation 12
3.9. Inward Growth and Configuration Complexity 13
3.10. Architectural Closure and the End of Universal Routability 13
3.11. Historical Context: Architectural Closure (1972-1994) . . 13

4. Observed Adaptive Responses 14
4.1. Relay-Centered Connectivity 14
4.2. Protocol Encapsulation and Substrate Reuse 15
4.3. Stateful Traversal and Long-Lived Associations 15
4.4. Identity Elevation and Application-Scoped Authority . . . 15
4.5. Silent Failure Tolerance and Retry Semantics 16
4.6. Transport-Layer Repair Attempts: SCTP and QUIC 16
4.7. Application-Guided Path Selection and Cost Signaling . . 17
5. Persistence and Normalization of Compensation 18
6. Indicators: Structural Load and Constraint 18
7. Analysis: Compensatory Mechanisms as Evidence 19
8. Post-Desire Path: Three Signals of an Unresolved Architectural Shift 20
8.1. RFC 7288: Firewalls as a Persistent Feature Without Formal Architectural Status 20
8.2. RFC 5218: When Widely Deployed Is Not the Same as Structurally Sound 21
8.3. RFC 7305: The Consequence: Control Migrates to Layer 7 . 21
8.4. Synthesis 22
9. Implications of the Present Equilibrium 22
9.1. Present Equilibrium 22
9.2. What This Reconstruction Establishes 23
10. IANA Considerations 23
11. Security Considerations 23
12. Informative References 24
Author’s Address 26

1. Purpose and Scope

This document reconstructs how the Internet arrived at its present connectivity equilibrium by examining observable adaptations, constraints, and deferred decisions over time. It does not assign fault, advocate specific remedies, or propose new protocol mechanisms. Instead, it seeks to explain why the system evolved as it did, given the pressures it faced and the locally rational responses available to its participants.

The analysis adopts a retrospective, systems-oriented perspective. It treats historical adaptations as evidence of underlying structural conditions rather than as errors or oversights. Decisions are evaluated in the context in which they were made, with attention to urgency, uncertainty, and available alternatives at the time. This framing is intentionally descriptive rather than corrective.

A central premise of this document is that systemic outcomes cannot be understood solely by examining individual design choices in isolation. Instead, they emerge from the interaction of multiple pressures, operating at different timescales, that shape what kinds of decisions are feasible, visible, or deferrable. The intent here is to surface those interactions.

This document is analytical rather than prescriptive. Its purpose is to make visible a pattern of systemic behavior that is otherwise easy to overlook precisely because the system has continued to function.

A companion document revisits end-to-end reasoning under these contemporary conditions and examines possible architectural response space. The present document confines itself to reconstruction and classification and does not propose remedies.

1.1. Central and Subsidiary Theses

1.1.1. Central Thesis

The Internet’s current connectivity equilibrium does not arise from the failure of a single architectural principle or protocol. Rather, it reflects the convergence of multiple eroded assumptions about physics, topology, authority, cost, and trust that once made ambient end-to-end connectivity inexpensive. As those assumptions eroded independently under new physical and policy constraints, the system responded by introducing mediation, buffering, and policy at multiple layers. The resulting equilibrium is stable not because the original assumptions still hold, but because compensatory mechanisms successfully absorbed their loss.

1.1.2. Subsidiary Thesis

Debates that localized the end-to-end problem primarily at the transport layer were not incorrect in their observations, but were constrained in scope by the urgency and visibility of transport-layer failures. They implicitly assumed that L4 was the first or only layer at which end-to-end semantics were withdrawn. In reality, analogous withdrawals had already occurred at the physical, link, and network layers, each for the same underlying reason: preventing a single participant from imposing unbounded cost on others. Structural pressures above and below the transport layer both demanded immediate attention and obscured the gradual loss of semantic clarity at L4, delaying focused reconsideration.

1.1.3. Scope Clarification

This observation is not intended to dismiss transport-layer research or to suggest that such work was conceptually misguided. Rather, it reflects the practical reality that urgent, layer-local failures necessarily shaped the framing of contemporaneous debate. Narrow focus under operational pressure should be understood as a constraint on visibility, not as an architectural error.

1.1.4. Clarifying Observation (Ambient Endpoints)

Throughout the stack, endpoints are ambient: each layer defines its own notion of an endpoint that is assumed to exist prior to higher-layer interaction. Physical endpoints exist as attached transceivers; link-layer endpoints exist as members of a broadcast or multicast domain; network-layer endpoints exist as addressable nodes within a routing scope; transport-layer endpoints exist as sockets and flows; and application endpoints exist as semantic actors.

End-to-end reasoning therefore depends on the continued ambient availability of endpoints at each layer. As mediation and scoping were introduced to contain cost and enforce policy, the ambient nature of endpoints was progressively withdrawn or made conditional at multiple layers. A recurring structural pressure underlying these changes was the need to prevent any single participant from imposing unbounded cost on others, whether through fault, misconfiguration, or asymmetric resource consumption. As ambient participation was withdrawn to bound such costs, higher layers were forced to compensate, doing so as effectively as possible using the authority and visibility available to them. This observation explains why end-to-end behavior degraded independently across layers without any single point of failure.

1.2. Ambient Endpoints and Their Progressive Withdrawal (By Layer)

* Physical Layer (L1): Attachment as Participation
  - Ambient assumption: If a device is physically attached, it can participate in communication on equal terms.
  - Withdrawal: Red/blue separation, switched media, and link termination replaced shared energy with bounded fault domains.
  - Pressure: A single faulty or malicious transmitter could impose unbounded disruption on all others.
  - Result: Physical attachment no longer implies ambient participation; existence becomes conditional and mediated.
* Link Layer (L2): Membership as Reachability

- Ambient assumption: Membership in a broadcast domain implies mutual reachability.
- Withdrawal: VLANs, multicast filtering, and suppression replaced flat broadcast with administratively scoped domains.
- Pressure: Broadcast amplification and heterogeneous media made shared fate expensive.
- Result: Link-layer endpoints remain, but membership is policy-defined rather than ambient.

* Network Layer (L3): Addressability as Existence

- Ambient assumption: An address implies routability and potential reachability.
- Withdrawal: Policy routing, routing-domain separation, and later firewalls conditioned reachability.
- Pressure: Divergent trust domains and administrative scale.
- Result: Addressability no longer implies permission or path availability.

* Transport Layer (L4): Packet Arrival as Conversation

- Ambient assumption: If packets arrive, a conversation may proceed; failure is explicit.
- Withdrawal: Admission control, state limits, silent drops, and middlebox mediation.
- Pressure: State exhaustion, asymmetric resources, and ambiguity of silence.
- Result: Transport endpoints persist, but conversational availability becomes inferred.

* Application/Semantic Layer (L7): Success as Correctness

- Ambient assumption: Successful interaction implies semantic correctness.
- Withdrawal: Retries, gateways, relays, and masking introduced ambiguity.
- Pressure: Uptime expectations and partial failure tolerance.
- Result: Semantic endpoints remain, but correctness is increasingly inferred.

This inventory provides the analytical baseline for the remainder of this document. Later sections treat these progressive withdrawals as observed structural conditions, not as isolated design mistakes.

2. Baseline Assumptions and Early Operating Conditions

Early Internet architecture assumed relatively stable hosts, cooperative administration, and ambient reachability. Hosts were institutionally operated, and participation implied adherence to shared norms and oversight.

Under these conditions, admission control and exposure were host-local concerns. Semantic authority, policy authority, and operational responsibility were closely aligned.

These assumptions reflected lived operational reality at the time and were sufficient for the Internet’s formative scale and threat model.

2.1. Early RFC Evidence, Grouped by Theme

The following historical material is drawn from early RFCs and related meeting notes. These sources are grouped thematically rather than chronologically in order to highlight recurring problem framings and system pressures that were recognized while the network was still forming. None of these documents should be read as definitive blueprints for later architecture; instead, they record how designers and operators understood emerging constraints in real time.

2.1.1. Mediation, Local Control, and Administrative Boundaries

Several early documents frame network interaction as mediated negotiation between autonomous systems, rather than as transparent end-to-end exchange.

* RFC 8 (1969) [RFC8] presents interaction as a sequence of steps across local control components: a user program establishes local arrangements, reaches a remote system, and requests service from that system’s own control program. This actor/system framing emphasizes locality and administrative authority over abstraction layering.

* RFC 706 (1975) [RFC706] explicitly proposes selective refusal of traffic at the Host/IMP boundary, allowing a host to instruct the network to discard messages from misbehaving or unwanted sources as early as possible. This reflects early recognition that unconditional acceptance is unsustainable and that refusal must occur at a control boundary.

Together, these sources show that mediation and refusal were treated as foundational capabilities, not as later security add-ons.

2.1.2. Identity, Accountability, and the Meaning of "Free"

Early discussions consistently treat network endpoints as accountable identities rather than anonymous communication primitives.

* RFC 147 (1971) [RFC147] defines sockets primarily as unique identifiers bound to processes and hosts, with explicit attention to logging and accounting. Communication is from one identifiable socket to another, reinforcing the notion of accountable endpoints.

* RFC 491 (1973) [RFC491] challenges the assumption that "free" network services must be loginless. Padlipsky argues that identity binding via login may still be required for authentication and access control, and proposes uniform free accounts as a portability compromise. This highlights early tension between convenience and semantic integrity.

These discussions anticipate later concerns about identity, attribution, and consent, and reject the idea that free services imply absence of control.

2.1.3. Fragmentation, Heterogeneous Environments, and Why "Normal" Features Were Deferred

Plurality and heterogeneity were recognized as intrinsic conditions from the outset, and early operational reality shaped which features were urgent.

* RFC 169 (1971) [RFC169] notes that the number of networks had already grown to the point where participants could not all be familiar with each other, and explicitly invites discussion of diverse systems, protocols, and user communities. Fragmentation is treated as a given, not as a deviation.

* RFC 898 (1984) [RFC898] reflects mature experience with heterogeneous gateways, subnetworks, and autonomous systems, documenting how routing, translation, and management complexity scale with diversity.

A related historical point is that many "normal" features associated with managed local networks, such as automatic configuration, routine endpoint discovery, and pervasive service location, were not treated as architectural necessities in the early Internet. This was not because such features were unknown, but because the environment did not yet demand them: early internetworking connected a relatively small number of large, institutionally operated hosts across administrative boundaries, rather than dense intranets of frequently

rebooting, mobile endpoints. In that setting, explicit local arrangements, operator knowledge, and manually coordinated configuration were sufficient, and the architectural forcing function was inter-networking between distinct domains rather than internal plug-and-play convenience.

As the Internet later grew inward into campuses and enterprises, accumulating large multi-LAN environments, higher endpoint churn, and widespread non-expert operation, automatic configuration and discovery became economically and operationally necessary, and the absence of first-class primitives increasingly had to be compensated elsewhere. RFC 1029 (1988) provides a concrete example of this inward growth pressure, addressing ARP scaling, bridge intelligence, reboot detection, and cache coherence in large multi-LAN Ethernet environments where frequent host churn and internal topology complexity had become dominant concerns.

2.1.4. Physical Reality, Delay, and Layer Blurring

Several early documents show that physical constraints immediately stress interaction models and blur later conceptual layer boundaries.

* RFC 263 (1971) [RFC263] describes a "very distant" Host/IMP interface in which the host participates directly in framing, CRC generation, and retransmission. Reliability and framing are treated as boundary-of-control concerns rather than as cleanly separated layers.

* RFC 346 (1972) [RFC346] observes that satellite delay renders character-at-a-time remote echo marginal or unusable, even when throughput is unchanged. Postel emphasizes buffering strategy and suggests relocating input/echo semantics closer to the user system.

These documents illustrate that delay and physical distance expose semantic assumptions early, forcing pragmatic integration across what would later be labeled layers.

2.1.5. Cost, Noise, Control-Plane Externalities, and the Turn Toward Managed High-Bandwidth Networks

Economic cost, background traffic, and control-plane scaling pressures appear early and intensify as bandwidth increases.

* RFC 392 (1972) [RFC392] measures host CPU and paging costs for network transmission, showing that the cost of moving data can exceed the cost of remote computation. Networking is treated explicitly as a distributed-systems cost problem rather than a free transport service.

* RFC 425 (1972) [RFC425] identifies "random prodding and poking" (e.g., host surveys) as a significant and unattributed source of overhead, and proposes consolidation and consent as remedies-an early recognition of background noise as a systemic externality.

* RFC 898 (1984) [RFC898] documents routing update storms, neighbor-probe scaling (e.g., N-squared behavior), and buffer exhaustion in gateways, illustrating how control-plane traffic can dominate useful forwarding work.

* RFC 1077 (1988) [RFC1077] synthesizes these concerns in the context of gigabit networking, explicitly reframing the future Internet as a management architecture. Importantly, this was not a speculative or marginal position: RFC 1077 reports the outcome of a DARPA-convened working group composed of principal architects, operators, and major stakeholders from the military, government, and research communities. It reflects the operational priorities of organizations that were already among the largest and most demanding users of packet networks, particularly in command-and-control, scientific computing, and secure communications contexts.

* RFC 1093 (1989) [RFC1093] makes the architectural consequences of these pressures operational. In defining the NSFNET routing architecture, it explicitly enforces policy-based filtering between the NSFNET backbone, regional networks, and peer networks such as ARPANET/MILNET. Certain routes are deliberately suppressed, metrics are reconstituted centrally, and trust is assigned by Autonomous System rather than by reachability alone. This document is notable as one of the earliest points where the evolving model of the Internet is acknowledged implicitly through implementation: the architects were aware that pure reachability was no longer sufficient, and encoded governance, policy, and functional separation directly into the routing fabric because they had to make the system operate at scale.

Taken together, these sources show a clear progression: increasing bandwidth does not eliminate cost or noise, but instead shifts the limiting factors toward control, coordination, security, governance, and explicit policy enforcement.

3. Emergence of Existential Stressors

The progressive withdrawal of ambient endpoints described earlier did not occur in a vacuum. It was driven by a set of existential stressors that demanded immediate response and shaped which adaptations were feasible, visible, or deferrable. These stressors were recognized early and recur throughout the historical record.

3.1. Fragmentation and Administrative Plurality

As documented as early as RFC 169 [RFC169], the network rapidly evolved into an environment of multiple, independently administered systems. Designers no longer assumed global familiarity, uniform policy, or shared objectives. This plurality forced early attention to gateway design, routing boundaries, and management coordination, and made purely uniform solutions impractical.

3.2. Physical Distance, Delay, and Interaction Breakdown

Physical realities such as propagation delay exposed fragile interaction semantics almost immediately. RFC 346 [RFC346] shows that even modest increases in delay (e.g., via satellite links) could render character-at-a-time interaction unusable, prompting discussion of buffering strategies and relocation of input/echo processing. These effects occurred well before Internet-scale deployment.

3.3. Cost and Host Resource Exhaustion

Economic viability emerged as a dominant constraint. RFC 392 [RFC392] demonstrates that host CPU time, paging behavior, and operating-system abstractions could make network transmission more expensive than remote execution itself. This reframed networking as a distributed-systems cost problem rather than a mere communications issue.

3.4. Background Traffic and Unattributed Load

Control-plane and exploratory traffic quickly became a measurable burden. RFC 425 [RFC425] documents how host surveys and other unsolicited probes generated significant overhead without clear attribution, motivating proposals for consolidation and explicit consent. These concerns foreshadow later issues with background chatter and steady-state coordination traffic.

3.5. Unconditional Acceptance and Denial of Service

The assumption that hosts must accept all traffic proved untenable. RFC 706 explicitly identifies denial-of-service risks from misbehaving peers and proposes selective refusal at the Host/IMP boundary. This represents early recognition that availability requires the ability to decline traffic before host resources are consumed.

3.6. Routing Scale, Control-Plane Costs, and Exit-Gateway Geometry

By the early 1980s, routing itself had become a stressor. RFC 898 [RFC898] documents how routing update floods, neighbor probing, and limited buffers strained gateways, and how thinking in terms of entrance and exit gateways reshaped autonomous systems into transit fabrics. These dynamics parallel later experiences with relay-centric architectures at higher layers.

3.7. Security Normalization: Routing Withdrawal, Filtering, and Firewalls

By the early 1990s, operational security controls such as routing withdrawal, packet filtering, and firewall choke points were no longer exceptional mechanisms but standard operational practice. RFC 1244 (Site Security Handbook) [RFC1244] treats these mechanisms as routine tools available to site operators, including selective route suppression, gateway filtering, and controlled connectivity.

A key inflection point for this normalization was the 1988 Internet worm. RFC 1135 (1989) [RFC1135], a retrospective on the incident, contains a blunt assessment in its Security Considerations: "If security considerations had not been so widely ignored in the Internet, this memo would not have been possible." In the aftermath, many sites tightened access, some disconnected entirely, and the community accelerated incident response coordination and perimeter controls.

3.8. Evolving Internet Membership: From IP Reachability to Application-Level Participation

RFC 1287 (1991) [RFC1287] makes explicit that the original IP-connectivity definition of the Internet had already broken down. Systems could be considered part of the Internet despite partial connectivity, policy filtering, or lack of IP reachability, so long as they participated at higher layers (e.g., RFC 822 mail). The architects proposed shifting the organizing principle of the Internet from IP addressability to application-level naming and directories.

3.9. Inward Growth and Configuration Complexity

RFC 1029 (1988) [RFC1029] documents the operational pressures that arise as the Internet grows inward into large multi-LAN environments: address resolution scaling, bridge intelligence, reboot detection, and cache coherence. This reinforces that partial visibility and constrained reachability can be expected outcomes of internal complexity and churn.

3.10. Architectural Closure and the End of Universal Routability

By the late 1980s and early 1990s, the Internet’s core architectural tensions were no longer latent. They were explicitly identified, debated, and-in key places-encoded into operational practice.

RFC 1093 (1989) [RFC1093] provides a concrete example of functional separation and policy-mediated reachability at backbone scale: military-only routes (ARPANET/MILNET) were deliberately suppressed from civilian regional backbones, with Autonomous Systems serving as trust and policy boundaries.

RFC 1627 (1994) "Network 10 Considered Harmful" [RFC1627], marks a clear self-awareness moment: the community recognized that the fully routable, globally unique IPv4 Internet was becoming operationally fragile under address exhaustion and policy constraints. While the specific compensations adopted later differed from what many hoped (e.g., NAT and application-layer identity became structural), the underlying pressures were already visible and the direction of travel was clear.

Taken together, these stressors explain why compensatory mechanisms emerged and hardened. They also show that many pressures commonly attributed to later Internet growth were visible-and actively discussed-by no later than the early 1990s.

3.11. Historical Context: Architectural Closure (1972-1994)

This history should not be read as a failure narrative. The record indicates that by the early 1990s the Internet’s core architectural tensions were already clearly identified and, in key operational networks, treated as constraints that could not be wished away.

Across the sources reviewed here, a consistent arc is visible:

* 1972-1975: Delay, background traffic, and selective refusal were already recognized as systemic issues (e.g., satellite delay effects, survey "noise", and early refusal/filtering proposals).

* 1984: Routing and gateway complexity, update scaling, and the inevitability of policy and control-plane costs were discussed as operational realities.

* 1988-1989: High-bandwidth planning reframed the Internet as a management architecture, while backbone routing explicitly enforced administrative separation and policy filtering.

* 1991: Security controls (withdrawal, filtering, firewalls) were normalized as routine operations, and the definition of "on the Internet" began shifting upward from IP reachability toward application-level naming, directories, and relay-mediated participation.

* 1994: The end of universal routability under IPv4 was recognized as a practical inflection point; subsequent decades largely operationalized compensations rather than discovering new categories of constraint.

This framing is essential context for revisiting end-to-end reasoning in a world where reachability is conditional, identities are increasingly application-scoped, and intermediaries are structural.

4. Observed Adaptive Responses

The adaptive responses that emerged as ambient reachability was progressively withdrawn can be grouped into several recurring patterns.

This section marks the transition from historical reconstruction to structural observation: these patterns are treated as convergent adaptations to shared constraints, not as a protocol-by-protocol survey.

These patterns appeared independently across applications, vendors, and administrative domains, yet converged on similar structural solutions.

4.1. Relay-Centered Connectivity

One of the earliest and most persistent adaptations was the introduction of relays. Rather than assuming that two endpoints could establish direct communication, systems increasingly routed interaction through one or more intermediary nodes that were known to be reachable from both sides.

Mail transfer agents, application-layer gateways, TURN-like relays, rendezvous servers, and later cloud-hosted service front ends all exemplify this pattern. Relays provided a point of policy enforcement, buffering, identity translation, and fault isolation. While they increased latency and centralized load, they dramatically reduced the requirement for mutual ambient reachability.

4.2. Protocol Encapsulation and Substrate Reuse

Another major adaptation was the reuse of widely permitted substrates to carry new application semantics. HTTP emerged as the dominant example of this pattern.

As early as RFC 3205 (2002) [BCP56], the IETF recognized that protocol designers were deliberately layering new services over HTTP in order to traverse firewalls, proxies, and network address translators. This practice was sufficiently widespread to require formal guidance, resulting in BCP 56. Two decades later, the same BCP was revised and reissued as RFC 9205 (2022) [BCP56], reflecting accumulated operational experience rather than a change in direction.

The persistence of BCP 56 over twenty years demonstrates that HTTP substrate reuse was not a transient workaround but a durable response to structural connectivity constraints.

4.3. Stateful Traversal and Long-Lived Associations

Where direct inbound reachability was unavailable, systems shifted toward models that established outbound-initiated, long-lived associations. These associations inverted the direction of connectivity: endpoints that could not accept unsolicited inbound traffic instead maintained persistent outbound sessions to rendezvous points.

Examples include message polling, push-notification channels, long-polling, WebSockets, and later QUIC-based connections. These techniques transformed connectivity from a stateless addressing problem into a stateful session management problem, trading simplicity for reliability under constrained reachability.

4.4. Identity Elevation and Application-Scoped Authority

As network-layer identity became unreliable or ambiguous, applications increasingly bound identity and authority at higher semantic layers. Authentication tokens, application-level identifiers, and service-specific namespaces replaced implicit trust in source addresses.

This shift aligned authority with mechanisms that applications could control, but further decoupled application semantics from network topology. Endpoints were no longer defined primarily by where they were located, but by what credentials or context they presented.

4.5. Silent Failure Tolerance and Retry Semantics

As ambient reachability became unreliable, applications adapted by treating silence as an expected condition rather than as an exceptional failure. Packet loss, filtering, middlebox interference, and policy-based drops are often indistinguishable from delay or congestion at the application layer.

Rather than assuming explicit failure signaling, applications adopted retry loops, timeouts, exponential backoff, and idempotent operations. These techniques allow progress in the presence of partial failure but shift complexity upward: correctness becomes probabilistic and inferred rather than explicit.

This adaptation increases robustness under constrained reachability but also obscures failure causes and complicates diagnosis. Silent tolerance trades semantic clarity for survivability, reinforcing the broader trend of compensating at higher layers for withdrawn ambient guarantees below.

4.6. Transport-Layer Repair Attempts: SCTP and QUIC

The Stream Control Transmission Protocol (SCTP) [RFC4960] represents an early attempt to preserve transport-layer semantic clarity in the face of eroding endpoint assumptions. Standardized around 2000, SCTP introduced multi-homing, association-based identity, path-aware failure detection, message framing, and multistreaming. Together, these features explicitly rejected the assumption that a single IP address uniquely and stably identifies a transport endpoint.

SCTP distinguished between path failure and peer failure, attempted to maintain semantic precision under partial failure, and treated transport associations, not addresses, as the primary unit of identity. In doing so, SCTP anticipated many later concerns about mobility, multihoming, and ambiguous silence.

However, SCTP assumed that new transport semantics could deploy transparently through the network. By the time of its standardization, that assumption had already been withdrawn: middleboxes, firewalls, and NATs were pervasive, and unfamiliar transport protocols were routinely blocked. As a result, SCTP’s technically sound repairs were largely displaced by compensations implemented above the transport layer.

QUIC [RFC9000], by contrast, represents a later and more successful adaptation. Rather than repairing L4 in place, QUIC relocates transport semantics into user space and runs over UDP, a substrate already widely permitted. QUIC encrypts most transport headers, preventing ossification by intermediaries, and treats connection identity, path migration, and congestion control as application-visible concerns.

The contrast between SCTP and QUIC is illustrative. SCTP attempted to restore ambient transport semantics that the network no longer supported. QUIC accepts mediation as structural and adapts by shifting authority upward, aligning deployment reality with semantic control. This contrast reinforces the broader pattern observed throughout this document: when ambient assumptions are withdrawn at a given layer, durable solutions tend to emerge by relocating responsibility rather than by attempting restoration in place.

4.7. Application-Guided Path Selection and Cost Signaling

A later and more explicit form of semantic elevation appears in the Application-Layer Traffic Optimization (ALTO) protocol (RFC 7285) [RFC7285]. ALTO exposed network cost, locality, and preference information as an application-consumable service, allowing endpoints to make informed choices among multiple reachable peers or resources.

This represented a qualitative shift in responsibility. Traditional routing determines how packets flow once a destination is chosen; ALTO assisted applications in deciding which destinations should be chosen in the first place. In effect, ALTO performed a form of quasi-source routing at L7: the network supplied advisory cost information, but the application selected targets and thereby shaped traffic patterns.

Cost, congestion, policy, and locality, once implicit properties of the network fabric, were surfaced explicitly to applications. This shift acknowledged that reachability alone no longer provided sufficient semantic guidance for efficient or stable behavior at scale.

ALTO did not replace routing, nor did it alter forwarding behavior. Instead, it compensated for the loss of ambient semantic information by elevating selected network knowledge to a controlled, advisory interface.

In practice, however, ALTO saw limited deployment outside a small number of research and operator-driven environments. Much like SCTP at the transport layer, it represented a semantically well-founded architectural repair that failed to align with prevailing deployment

incentives. Application developers largely bypassed ALTO in favor of self-managed heuristics, static configuration, or embedding cost and locality inference directly into application logic, often using widely permitted substrates and measurement-based adaptation.

As a result, ALTO functions primarily as evidence of architectural recognition rather than as a dominant operational mechanism: it demonstrates that the need for explicit cost and locality signaling was understood, even as most implementations chose compensatory approaches that avoided new dependencies on network-provided control planes.

5. Persistence and Normalization of Compensation

Over time, compensatory mechanisms ceased to be exceptional. What began as fallback behavior hardened into steady-state infrastructure. Relay paths became primary paths, and indirect connectivity became the default assumption rather than the contingency plan.

This persistence had several reinforcing effects. First, widespread deployment increased the return on further investment in compensatory mechanisms, making them more capable and more attractive. Second, their effectiveness reduced the frequency of visible failures that might have triggered architectural reconsideration.

In the presence of more urgent, existential concerns, other issues were routinely deferred until they themselves became urgent. Because compensatory mechanisms continued to work, the cost of revisiting underlying assumptions appeared higher than the cost of continued adaptation.

As a result, the system accumulated technical and conceptual debt without a clear moment at which repayment appeared necessary or even desirable.

When a system model depicts a viable path that is consistently avoided, the discrepancy should be attributed to the model or the path, not to the actors responding rationally to observed constraints.

6. Indicators: Structural Load and Constraint

Despite continued operation, the system began to exhibit recurrent indicators of underlying load and constraint. These indicators were not catastrophic failures, but patterns that suggested increasing reliance on compensation and diminishing alignment between architectural assumptions and operational reality.

Such indicators included loss of locality, concentration of load onto shared infrastructure, opaque or delayed failure modes, and growing difficulty in determining where authority and responsibility for communication decisions actually resided.

These signals were often diffuse and probabilistic rather than binary. They manifested as degraded efficiency, increased complexity, or brittleness under stress rather than as immediate outages. Because the system continued to function, they were tolerated rather than treated as forcing events.

The absence of a single, unambiguous failure made it difficult to justify a coordinated architectural response.

7. Analysis: Compensatory Mechanisms as Evidence

When a system model depicts a viable path that is consistently avoided, the discrepancy should be attributed to the model or the path, not to the actors responding rationally to observed constraints.

A familiar example is the formation of pedestrian "desire paths." Such paths arise when users repeatedly choose routes that better reflect actual needs than those anticipated by the original design. Over time, repeated use alters the environment itself, and what began as an exception becomes a structural feature.

ALTO illustrates an attempt to formalize application-visible cost signaling after routing and admission authority had already moved. Its limited impact is therefore informative: it demonstrates both the recognition of the problem and the difficulty of addressing it once compensatory mechanisms have become structural.

In the Internet’s case, compensatory connectivity mechanisms functioned as desire paths. They revealed a mismatch between architectural assumptions about reachability and the operational conditions under which the system was actually used. Their persistence and success transformed them from temporary adaptations into defining characteristics of the system.

Seen in this light, compensatory mechanisms are not merely technical artifacts; they are empirical signals about where system models no longer align with reality.

A similar interpretive stance appears in human-system design. When users repeatedly avoid an architected path, analysis treats the avoidance as evidence of misaligned assumptions rather than as user error. Norman’s discussion of "desire paths" frames such behavior as

empirical data about real constraints and incentives, not as deviation from intent [Design]. The persistence and convergence of compensatory mechanisms in Internet connectivity can be understood in the same way: not as architectural failure, but as evidence that certain assumptions no longer held under operational conditions.

8. Post-Desire Path: Three Signals of an Unresolved Architectural Shift

The desire-path argument establishes that persistent operator behaviour is evidence of a mismatch between the model and the environment. The following RFCs are useful precisely because they show the Internet recognizing the mismatch while stopping short of formally resolving it.

The observations in this section are descriptive rather than prescriptive: they examine how the mismatch has been acknowledged and framed, not how it ought to be resolved.

8.1. RFC 7288: Firewalls as a Persistent Feature Without Formal Architectural Status

RFC 7288 [RFC7288] is notable less for any specific proposal than for the careful position it occupies within the existing architectural narrative.

The document acknowledges the widespread and long-standing presence of firewalls, and does so in a pragmatic and operationally grounded way. At the same time, it deliberately avoids treating firewalls as a permanent structural element of the Internet architecture. Instead, they are discussed as policy-enforcing devices that exist alongside the architecture rather than within its formal core.

From a desire-path perspective, this restraint is understandable. RFC 7288 operates within an architectural framework that continues to value the end-to-end principle as a guiding ideal, even as practice has moved away from ambient inbound reachability. Rather than declaring that shift complete, the document treats firewalls as an external constraint that must be accommodated.

The consequence of this position is not denial, but deferral. Firewalls are assumed to be present in practice, yet their ubiquity is not elevated to a baseline architectural condition. Subsequent designs are therefore encouraged to cope with their existence rather than to integrate them as a first-class premise, leading to repeated work on traversal, discovery, and rendezvous mechanisms instead of an explicit acknowledgement that ambient inbound reachability is no longer the norm.

In this sense, the desire path is clearly visible, but the architectural map remains intentionally conservative about redrawing its boundaries.

8.2. RFC 5218: When Widely Deployed Is Not the Same as Structurally Sound

RFC 5218 [RFC5218] provides a useful corrective by explicitly cautioning against equating deployment success with architectural merit.

The Internet has repeatedly adopted mechanisms that were operationally expedient under pressure, such as address sharing, middleboxes, and application-layer workarounds, without those mechanisms being clean fits for the original architectural model. RFC 5218 recognizes that popularity can arise from necessity, inertia, or lack of alternatives, rather than from correctness.

This distinction matters here because the current connectivity equilibrium is often defended on the grounds that it works or is widely used. RFC 5218 reminds us that such arguments describe outcomes, not structure.

The desire-path framework explains why this happens. When the environment changes faster than the model, actors will choose survivable routes even if they deform the original plan. Over time, these routes harden, not because they are ideal, but because they are viable.

RFC 5218 gives us permission to say plainly that the Internet’s current shape may be stable without being architecturally resolved.

8.3. RFC 7305: The Consequence: Control Migrates to Layer 7

RFC 7305 [RFC7305] is best read as an observation about where meaningful decisions now occur.

As lower-layer assumptions about reachability, symmetry, and transparency eroded, applications were forced to compensate. Authentication, discovery, mobility, policy, and even routing intent increasingly moved upward, until application protocols became the only layer with sufficient context to function reliably.

The practical outcome is that many decisions traditionally associated with the network or transport layers are now made at layer 7, because only the application can see across NATs, firewalls, relays, and policy boundaries.

This is not a design choice so much as a consequence of earlier non-decisions. By declining to formally acknowledge the withdrawal of ambient end-to-end reachability, the architecture implicitly delegated responsibility upward.

The Internet still speaks in layers, but it now decides almost exclusively at the top.

8.4. Synthesis

Taken together, these RFCs describe a system that has adapted successfully while avoiding a full architectural reckoning.

* RFC 7288 shows a feature treated as temporary long after it became permanent.
* RFC 5218 warns against mistaking survival for correctness.
* RFC 7305 documents the resulting migration of control into application space.

The desire paths are visible, continuous, and rational. What remains unresolved is not whether the Internet has adapted, but whether its architecture has yet caught up with its own behaviour.

9. Implications of the Present Equilibrium

The reconstruction above yields both an observable system state and a set of limits on what can be inferred from that state. The following sections address these together: first by characterizing the present connectivity equilibrium as it exists, and then by clarifying what the reconstruction establishes about that equilibrium.

9.1. Present Equilibrium

The Internet has settled into an equilibrium defined by these accumulated adaptations. This equilibrium is stable under current constraints and has enabled continued growth, innovation, and deployment. It is not characterized by collapse or obvious dysfunction.

At the same time, this stability depends on the continued effectiveness of compensatory mechanisms. The system operates by routing around certain assumptions rather than revisiting them directly. As a result, architectural questions concerning endpoints, authority, and reachability are deferred rather than resolved.

From a systems perspective, this equilibrium resembles a metastable regime: locally stable and resilient to small perturbations, yet dependent on sustained compensation and lacking strong restoring forces should underlying conditions change.

9.2. What This Reconstruction Establishes

This reconstruction suggests that the present connectivity model is not the result of a single decision or omission, but of sustained rational deferral under pressure. Major existential concerns demanded immediate action; secondary misalignments were tolerated because they admitted local and effective compensation.

The historical record examined here is consistent with this pattern. The adaptations that preserved functionality also reshaped the system, making certain architectural questions harder to see precisely because they were successfully avoided.

The presence of a stable equilibrium should not be read as an endorsement of that equilibrium. Stability here denotes persistence under prevailing constraints, not architectural optimality or normative correctness.

This document does not establish that the present equilibrium is unstable, undesirable, or incorrect. It establishes only that the conditions which once justified deferring certain architectural questions have changed, making those questions newly visible.

This document does not propose remedies, evaluate counterfactual architectures, or predict future outcomes. Its contribution is to clarify how the Internet arrived at its current state, and why questions about the suitability of that equilibrium have only recently become visible again.

10. IANA Considerations

This document has no IANA actions.

11. Security Considerations

This document is purely descriptive and retrospective. It does not propose new protocols, mechanisms, procedures, or operational practices, nor does it recommend changes to existing ones.

As such, it introduces no new security considerations beyond those already present in the systems and practices discussed. Any security-relevant mechanisms referenced are included solely as historical and architectural context.

12. Informative References

[Design] Norman, D. A., "The Design of Everyday Things", New York, Basic Books; rev. ed., ISBN 978-0465050659, 2013.

[RFC8] Deloche, G., "ARPA Network Functional Specifications", RFC 8, DOI 10.17487/RFC0008, May 1969, <https://www.rfc-editor.org/info/rfc8>.

[RFC147] Winett, J., "Definition of a socket", RFC 147, DOI 10.17487/RFC0147, May 1971, <https://www.rfc-editor.org/info/rfc147>.

[RFC169] Crocker, S., "COMPUTER NETWORKS", RFC 169, DOI 10.17487/RFC0169, May 1971, <https://www.rfc-editor.org/info/rfc169>.

[RFC263] McKenzie, A., ""Very Distant" Host interface", RFC 263, DOI 10.17487/RFC0263, December 1971, <https://www.rfc-editor.org/info/rfc263>.

[RFC346] Postel, J., "Satellite Considerations", RFC 346, DOI 10.17487/RFC0346, May 1972, <https://www.rfc-editor.org/info/rfc346>.

[RFC392] Hicks, G. and B. Wessler, "Measurement of host costs for transmitting network data", RFC 392, DOI 10.17487/RFC0392, September 1972, <https://www.rfc-editor.org/info/rfc392>.

[RFC425] Bressler, R., ""But my NCP costs $500 a day"", RFC 425, DOI 10.17487/RFC0425, December 1972, <https://www.rfc-editor.org/info/rfc425>.

[RFC491] Padlipsky, M., "What is "Free"?", RFC 491, DOI 10.17487/RFC0491, April 1973, <https://www.rfc-editor.org/info/rfc491>.

[RFC706] Postel, J., "On the junk mail problem", RFC 706, DOI 10.17487/RFC0706, November 1975, <https://www.rfc-editor.org/info/rfc706>.

[RFC898] Hinden, R., Postel, J., Muuss, M., and J. Reynolds, "Gateway special interest group meeting notes", RFC 898, DOI 10.17487/RFC0898, April 1984, <https://www.rfc-editor.org/info/rfc898>.

[RFC1029] Parr, G., "More fault tolerant approach to address resolution for a Multi-LAN system of Ethernets", RFC 1029, DOI 10.17487/RFC1029, May 1988, <https://www.rfc-editor.org/info/rfc1029>.

[RFC1077] Leiner, B., "Critical issues in high bandwidth networking", RFC 1077, DOI 10.17487/RFC1077, November 1988, <https://www.rfc-editor.org/info/rfc1077>.

[RFC1093] Braun, H., "NSFNET routing architecture", RFC 1093, DOI 10.17487/RFC1093, February 1989, <https://www.rfc-editor.org/info/rfc1093>.

[RFC1135] Reynolds, J., "Helminthiasis of the Internet", RFC 1135, DOI 10.17487/RFC1135, December 1989, <https://www.rfc-editor.org/info/rfc1135>.

[RFC1244] Holbrook, J. and J. Reynolds, "Site Security Handbook", RFC 1244, DOI 10.17487/RFC1244, July 1991, <https://www.rfc-editor.org/info/rfc1244>.

[RFC1287] Clark, D., Chapin, L., Cerf, V., Braden, R., and R. Hobby, "Towards the Future Internet Architecture", RFC 1287, DOI 10.17487/RFC1287, December 1991, <https://www.rfc-editor.org/info/rfc1287>.

[RFC1627] Lear, E., Fair, E., Crocker, D., and T. Kessler, "Network 10 Considered Harmful (Some Practices Shouldn’t be Codified)", RFC 1627, DOI 10.17487/RFC1627, July 1994, <https://www.rfc-editor.org/info/rfc1627>.

[RFC4960] Stewart, R., Ed., "Stream Control Transmission Protocol", RFC 4960, DOI 10.17487/RFC4960, September 2007, <https://www.rfc-editor.org/info/rfc4960>.

[RFC5218] Thaler, D. and B. Aboba, "What Makes for a Successful Protocol?", RFC 5218, DOI 10.17487/RFC5218, July 2008, <https://www.rfc-editor.org/info/rfc5218>.

[RFC7285] Alimi, R., Ed., Penno, R., Ed., Yang, Y., Ed., Kiesel, S., Previdi, S., Roome, W., Shalunov, S., and R. Woundy, "Application-Layer Traffic Optimization (ALTO) Protocol", RFC 7285, DOI 10.17487/RFC7285, September 2014, <https://www.rfc-editor.org/info/rfc7285>.

[RFC7288] Thaler, D., "Reflections on Host Firewalls", RFC 7288, DOI 10.17487/RFC7288, June 2014, <https://www.rfc-editor.org/info/rfc7288>.

[RFC7305] Lear, E., Ed., "Report from the IAB Workshop on Internet Technology Adoption and Transition (ITAT)", RFC 7305, DOI 10.17487/RFC7305, July 2014, <https://www.rfc-editor.org/info/rfc7305>.

[RFC9000] Iyengar, J., Ed. and M. Thomson, Ed., "QUIC: A UDP-Based Multiplexed and Secure Transport", RFC 9000, DOI 10.17487/RFC9000, May 2021, <https://www.rfc-editor.org/info/rfc9000>.

[BCP56] Best Current Practice 56, <https://www.rfc-editor.org/info/bcp56>. At the time of writing, this BCP comprises the following:

Nottingham, M., "Building Protocols with HTTP", BCP 56, RFC 9205, DOI 10.17487/RFC9205, June 2022, <https://www.rfc-editor.org/info/rfc9205>.

Author’s Address

Erik Fjeldstrom
Independent
Email: erik_fjeldstrom@yahoo.ca