去中心化聲譽網絡：
自然社會組織的框架

Pavel Kudrna
Prague, Czechia
Draft v1 — March 2026

Abstract

公義感——即相信錯誤會被糾正、功績會得到回報的信念——是社會最強大的凝聚力量。當中央化的治理機構因為強制執行一項權利的成本超過該權利本身的價值而無法提供這種公義感時，社會凝聚力便會侵蝕。現行的制度結構在系統性上未能解決相當一大類的糾紛，其失敗方式與中央計劃經濟未能有效配置資源的方式如出一轍：透過資訊不對稱、缺失的回饋循環，以及決策者與其決策後果之間的脫節。本文提出一個聲譽社交網絡（Reputation Social Network，RSN）：一個建基於去中心化識別碼（Decentralized Identifiers，DID）之上、去中心化、不可腐化、抗審查的通訊層，使化名參與者能夠發佈、驗證及取用關於現實世界行為的聲譽資訊。其核心機制建立於三項設計原則之上：（1）一種不對稱的成本結構，其中讀取聲譽資料成本低廉，而寫入則需要可驗證的付出；（2）一種基於一致性雜湊（consistent hashing）的非確定性驗證者選擇協議，透過不斷升級的迭代成本為激進主張定價；以及（3）一種市場驅動的聲譽權威模型，當中私營驗證者以其累積的聲譽為其所背書主張的準確性作押注。由此產生的架構在沒有中央協調的情況下催生出湧現式的精英制度（meritocracy），並使得從現行國家管理系統出發、循序漸進且可逆轉的遷移路徑成為可能。

1 引言

1.1 中央化信任的問題

現代治理建基於一項根本假設：中央化機構能夠大規模地在陌生人之間居間調解信任。法院裁決糾紛。登記處證明所有權。監管機構執行標準。這些機構設計於一個資訊稀缺、通訊緩慢、而將權力委託予中央機關是唯一可行的協調機制的年代。然而，中央化治理失敗的結構性原因與中央計劃經濟相同：資訊不對稱、缺失的回饋循環，以及決策者與其決策後果之間的脫節。

舉例而言，當機構調解的成本超過糾紛的價值時，這項假設便會失效。試想一位租客發現其租住的公寓缺乏法律要求的電力安全證明——這是一種對生命構成即時危險的狀況。租客在法律上撤銷合約的權利毫不含糊。然而，透過法院系統強制執行該項權利的成本，卻超過押金及應償還損失的金錢價值，即使計入潛在的法定賠償亦然 [1]。理性的回應是承受損失並離開。

這並非一個病態的邊緣個案。它是相當一大類糾紛的預設經驗——這些糾紛中的損害是真實的，但金錢利害關係卻低於制度性強制執行在經濟上變得理性的門檻。其結果是一個在結構上偏袒惡意行為者的系統：那些以低於此門檻的規模傷害他人的人，可以肆無忌憚地行事，因為尋求公義的成本落在受害一方身上。

上述例子取材自作者的法律環境——捷克的民法系統。在其他法律傳統中——以判例為本的普通法、習慣法、混合系統——公義會以不同的方式及不同的程度失效，視乎該司法管轄區的文化及程序特性而定。讀者很可能會從自己的處境中認出相應的例子。在結構上具普遍性的是這一模式：價值低於經濟理性強制執行門檻的糾紛，最終以受害一方承受損失告終。RSN 並不承諾完美的公義——它只是減少當制度性強制執行不符經濟效益時、惡意行為者所享有的結構性優勢。

1.2 為何現有方案有所不足

去中心化身份（DID）框架 [2] 為自我主權身份管理提供了密碼學機制，但主要聚焦於身份驗證，而未有處理行為聲譽這一更廣泛的問題。

靈魂綁定代幣（Soulbound Tokens，SBT） [3] 掌握了聲譽不可轉讓這一洞見，但依賴具有可擴展性限制的區塊鏈基礎設施，並且未有處理支配資訊錄入的經濟誘因。諸如功績代幣（例如 Liberland）等相關概念共享類似的理念，但仍受制於特定司法管轄區的框架。

去中心化自治組織（DAO） [4] 透過智能合約實行治理，但複製了金權政治的動態，其中影響力與資本成正比。二次方投票（Quadratic voting）[5] 部分緩解了這一點，但限制了實際採用。DAO 亦面臨根本的預言機問題（*oracle problem*）：智能合約在沒有可信外部來源的情況下無法可靠地驗證

現實世界的狀態，而這個問題在區塊鏈架構內並無通用的解決方案。

沒有任何現有系統能夠將去中心化、抗審查、化名性、可驗證的錄入成本，以及湧現式共識結合於一身。

1.3 貢獻

本文作出以下貢獻:

1. 定義**聲譽社交網絡（RSN）**，一個擴展 DID 框架以支援雙邊聲譽交換的去中心化協議。
2. 一種基於一致性雜湊的**非確定性驗證者選擇協議** [6]。
3. **昂貴激進主義原則（Expensive Radicalism Principle）**，透過三條複合的成本渠道，按主張與網絡共識的距離為其定價。
4. 一種具不對稱誘因的**可信權威（Reputable Authority）模型**，其中虛假背書的代價災難性地遠高於合法收費。
5. 一條透過並行財政基礎設施實現的**漸進遷移路徑**。

2 設計原則

RSN 架構受五項不可妥協的設計原則所約束。

延續性與演化。系統在一段不定長度的過渡時期內與現行機構共存。過渡在每一階段都必須是可逆轉的。

自願性與責任。參與是自願的，但自願性與責任相輔相成：一位承擔起某項制度職能控制權的參與者，同時承擔起隨之而來的義務。

多元性與文化中立。共識由雙邊互動湧現而生，而非由系統設計者強加的預設規則所產生。

韌性與抗審查。審查網絡的成本必須超過容忍它的成本。唯有徹底的極權主義——付出毀滅性的政治及經濟代價——才能壓制這個系統。

共識與執行。系統將人類趨向於小器、怨恨及報復性滿足的傾向納入，作為承重的特性。不當行為並非被禁止，而是被定價。

3 系統概覽

3.1 聲譽社交網絡

RSN 是一個去中心化的點對點通訊層，參與者於其中交換關於現實世界行為的可驗證資訊。其界定特性包括：去中心化且不可審查的基礎設施；透過 DID 進行的化名參與；不對稱的成本結構（讀取成本低廉，寫入成本高昂）；密碼學可驗證性；以及**標準支援**——為透過網絡傳播的資訊、為權威所提供的服務（主張組成、背書、見證服務）、以及為政策格式（包括評估交易對手聲譽及評估政策不符風險——即申報行為與實際行為之間差異——的規則）而設的機器可讀格式。

3.2 架構組件

RSN 由四個主要組件組成（圖 1）:

1. **DID 層。**帶有申報規則、聲譽元數據及網絡路由的參與者身份。
2. **主張協議。**用於發佈關於現實世界事件斷言的結構化格式。
3. **驗證網絡。**使用一致性雜湊指派驗證者的去中心化協議。
4. **聲譽聚合層。**產生人類可讀聲譽摘要的服務。

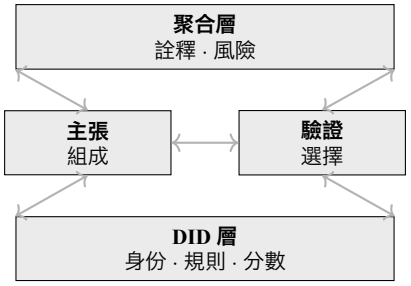

Figure 1: RSN 四層架構。

3.3 參與者角色

一宗驗證交易牽涉多達六種不同的 DID 角色（圖 2）：**發行者（Issuer）**（DID_I，創建並提交主張）、**對象（Subject）**（DID_S，主張所關於的 DID——可與發行者重疊）、**權威（Authority）**（DID_A，收費為主張背書；亦可提供見證服務——可選）、**見證者（Witness）**（$DID_{W_{1..n}}$，透過挑戰碼隱身監察驗證者誠信——可選）、**驗證者（Verifier）**（DID_V，經演算法選定以發佈主張），以及 **RSN**（已驗證主張獲發佈之網絡）。任何角色均可委託予另一 DID（第 7.4 節）。權威通常會將背書與見證服務捆綁在一起，但這兩項職能在邏輯上互相獨立。

3.4 不可腐化性：四種力量

RSN 的不可腐化性並非源自單一機制，而是源自四種同時運作的力量。任何一種單獨都不足夠；唯有它們的結合，才使腐化的嘗試在經濟上變得不理性。

不可預測的目標。每宗主張的驗證者透過一致性雜湊非確定性地選定（第 5.3 節）。攻擊者無法為賄賂而預先鎖定特定的驗證者，因為驗證者的身份是主張內容及先前迭代的函數，而非發行者選擇的函數。

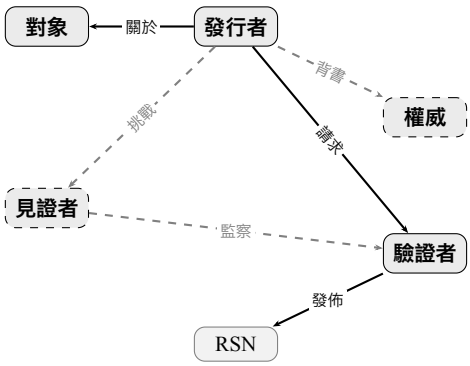

Figure 2: 驗證交易中的 DID 角色。虛線 = 可選（權威、見證者）。

聲譽槓桿——上升慢，下降快。聲譽只能透過持續一致的行為循序漸進地獲得。然而，它卻可以因一次欺詐性的背書而災難性地喪失（第 6.2 節）。這種不對稱意味著多年累積的聲譽可被一次不誠實的背書所摧毀；最優策略始終是拒絕可疑的主張。

公開承諾。每位 DID 持有者都公開申報其政策——一套他們承諾遵守的機器可讀規則。申報與實際行為之間的分歧是可偵測的，並被定價為一種比底層違規更嚴重的聲譽罪過（第 7.3 節）。因此，偽善比直接的不誠實代價更高。

網狀攻擊成本不對稱。審查或破壞網絡穩定的成本隨網絡的規模及密度非線性增長，而容忍它的成本則維持不變。要令審查有利可圖，一個中央化的行為者必須在整個網絡範圍內施加審查——所需的措施與任何可設想的收益完全不成比例（第 10 節）。

4 去中心化身份模型

4.1 具特殊屬性的 DID

RSN 擴展了 W3C DID Core 規範 [2]，加入：**申報規則**（機器可讀的行為承諾）、**聲譽元數據**（聚合的行為分數），以及**網絡路由**（與穩定的環形位置分開的可變洋蔥網關）。

4.2 主張生命週期

一宗主張經歷六個階段（圖 3）：組成、可選的權威背書、透過一致性雜湊進行驗證者選擇、驗證裁決（接受，或拒絕並迭代）、發佈，以及對所有各方的聲譽更新。

4.3 與 W3C DID Core 的關係

RSN 的身份模型是 W3C DID Core 規範 [2] 的一個真超集。所有 RSN DID 都是有效的 W3C DID。RSN 特有的擴展被編碼為在專屬 DID 方法規範中所定義的 DID 文檔屬性，與諸如 ION [7] 及 Ceramic [8] 等現有基礎設施兼容。

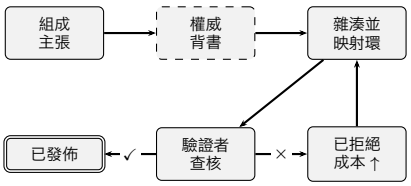

Figure 3: 帶迭代循環的主張生命週期。

5 資訊驗證與傳播

5.1 資訊錄入的成本

RSN 的核心經濟原則是讀取與寫入之間的不對稱。對於身處 DID 網絡之內、且擁有與其聲譽相稱之取用權的參與者而言，讀取聲譽資料的邊際成本趨近於零——新加入者必須逐步賺取更廣泛的取用權（見反吸血）。寫入——理解為聲譽持久地物化進入網絡，而非一次性的技術行為——需要跨越三個維度的可驗證累積投資：**時間**（花費於一致行為、履行承諾及培育關係的人生歲月）、**精力**（投入於誠實交易、關顧夥伴關係及長遠可信賴度的付出），以及**金錢**（對自己名譽的投資——專業發展、自身工作的質素、對所造成損害的賠償）。聲譽無法即時購得——它是終生累積的成果，而系統只是將其變得可見，並可跨情境轉移。相比這項底層投資，協議的一次性技術成本（密碼學簽名、驗證者費用）微不足道。

5.2 社交圖譜與聯絡深度

RSN 從根本上是一個社交網絡：每個 DID 增添聯絡人（其他允許此連接的 DID）。這些聯絡人有他們自己的聯絡人，如此類推。演算法的驗證者搜尋在可配置深度 h 的鏈接聯絡人範圍內運作（例如 $h = 3$：一個人的直接聯絡人、他們的聯絡人，以及聯絡人的聯絡人）。此架構並不需要單一的全球區塊鏈——它自然而然地有利於社群／叢集的湧現，並與其他社群有所重疊。每位 DID 參與者都必須有一份已發佈的**政策**——一套管治如何評估傳入請求的機器可讀規則。違反政策會作為負面的產物記錄於網絡之中。

5.3 演算法驗證者選擇

驗證協議採用一致性雜湊 [6]（圖 4）。驗證是一項收費服務：驗證者收費運作，締造一個由競爭驅動定價、速度及可靠性的市場。

步驟 1。主張文檔與上一次迭代的雜湊結果（或

首次迭代時的 ∅）連接並進行雜湊:

$$\text{pos} = f_h(\text{claim} \| \text{prev\_hash}) \tag{1}$$

演算法必須包含所有先前請求及回應的完整路徑——而不僅僅是上一次迭代的雜湊。每位驗證者的完整回應（接受、拒絕、報價、理由）都會附加到不斷增長的文檔中，並在每一步透過密碼學簽名可供驗證。

步驟 2。雜湊被映射到一致性雜湊環上的 N 個均勻分佈的位置（例如，當 $N = 4$: $+0°, +90°, +180°, +270°$）。從每個位置起，順時針搜尋選出最接近的 DID 作為驗證者候選人（圖 5）。N 是一個可變參數，可取決於當前活躍 DID 的百分比、一天中的時間，或網絡歷史上的回應速度。

步驟 3。驗證者接受（發佈，並押注聲譽）或拒絕（帶著拒絕雜湊返回步驟 1）。當某節點聲稱在線卻不作回應時，下一次請求必須包含所有 N 個先前驗證者候選人的所有回應。

反吸血。目標 DID 可為來自聲譽甚少的新 DID 的查詢設定費用或取用限制。此分級機制（而非二元）迫使新加入者在能自由取用他人資料之前，先透過真實活動建立聲譽。

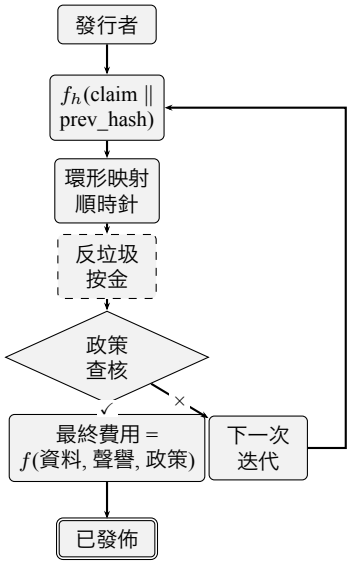

Figure 4: 驗證者選擇協議。反垃圾按金為可選，並可退還。最終費用僅在接受時支付。

每次拒絕都會將驗證者的完整回應（包括理由及條件）附加到主張文檔中，擴充其內容。從擴充後的文檔，非確定性地計算出一個新雜湊，映射到一個不同的環形位置並選出一位不同的驗證者。記錄完整路徑是強制性的——發行者無法預測或影響下一位驗證者。若某驗證者儘管有兼容的申報政策卻無視請求，見證者（如在場）會記錄此事，而發行者可在最終發佈時附上見證證據。

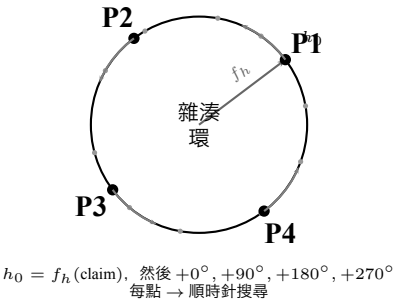

Figure 5: 多點選擇（N=4）。雜湊 h_0 設定偏移；由 h_0 均勻分佈 4 個點。

5.4 見證者協議

見證者角色透過一種密碼學挑戰碼協議，為驗證者的誠信提供隱身的問責:

步驟 W1。請求者簽署驗證請求，並將其發送予 N_w 位見證者——來自請求者社交網絡的聯絡人，或收費運作的專門見證服務供應商。

步驟 W2。每位見證者 w_i 簽署整份已簽署的請求,保留原始簽名 σ_i,並計算一個挑戰碼 $c_i = h(\sigma_i)$。挑戰碼被附加到請求上。

步驟 W3。現在攜帶 N_w 個挑戰碼的請求被發送予驗證者。驗證者觀察到這些碼，但無法確定見證者是誰，或這些碼是否對應真實的簽名。

步驟 W4（誠實的驗證者）。若驗證者按照其申報的政策作出回應，挑戰碼便維持不透明。見證者永不被揭露。不會發佈任何額外資料。

步驟 W5（違反政策）。若驗證者違反其政策（無視請求、回應前後不一），請求者持有見證者的原始簽名 σ_i。這些可作為佐證資料發佈：任何人都可驗證 $c_i = h(\sigma_i)$，證明見證者曾在場。請求者可獨立發佈——驗證者已在其回應中簽署了挑戰碼。

虛張聲勢屬性。請求者可以隨機值替換部分或全部挑戰碼。驗證者無法區分真正的碼（由高聲譽權威支持）與雜訊。這種不確定性正是執行機制: 每宗請求都帶有一位受尊敬的權威正在隱身觀察的風險。維持這種壓力的成本趨近於零（生成隨機數），而不誠實對驗證者而言的潛在代價卻是災難性的。即使在沒有實際見證者的情況下，誠實行為亦受到激勵。

權威作為見證者。為主張背書的權威可捆綁見證服務:「我為你的主張背書，並在驗證期間充當隱身見證者。」這是一種自然的商業模式——權威早已掌握情境。然而，這兩種角色在邏輯上互相獨立。

5.5 昂貴激進主義原則

三條成本渠道同時複合（圖 6）:

渠道 1：迭代成本。每次拒絕都迫使一次消耗時

間及資源的新迭代。線性增長。

渠道 2：文檔膨脹。每次迭代都將完整的驗證者回應加入主張文檔。增長是線性的，但帶有一個基底偏移：初始負載（主張內容、權威背書、密碼學簽名）已具有不容忽視的大小 B_0。k 次迭代後的總大小：$S(k) = B_0 + k \cdot \Delta$，其中 Δ 是平均回應大小。

渠道 3：驗證者風險溢價。風險是主觀的。驗證者評估請求方 DID 的申報政策是否與驗證者自身的商業、社會或政治利益相衝突。溢價可隨著與驗證者「理想」的感知距離而指數式升級：$P(d) = \alpha \cdot e^{\beta d}$，其中 d 是驗證者的主觀距離度量。超越個人的禁區邊界，驗證者可完全拒絕。

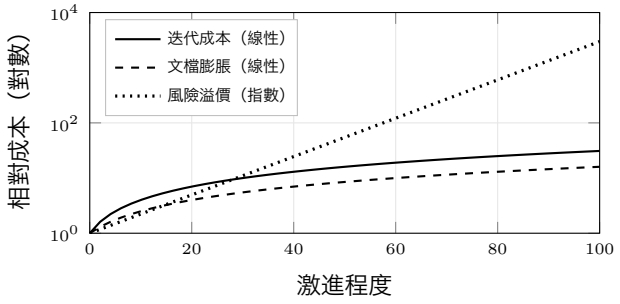

Figure 6: 三條渠道的成本升級。風險溢價在高激進程度下佔主導。

關鍵在於，這並非審查。沒有任何主張被禁止。系統為風險定價（表 1）。

Table 1: 各類主張類型的成本比較。

類型	迭代	文檔	費用	總計
可信	$k_{\min}$	B_0	$P_{\min}$	低
具爭議	k_{med}	$B_0 + k\Delta$	$\alpha e^{\beta d}$	中等
激進	$k_{\max}$	$\gg B_0$	$\gg P_{\min}$	極高
欺詐	$\to \infty$	—	—	無法發佈

6 聲譽與風險評估

6.1 聲譽累積模型

聲譽呈現三種屬性：**緩慢累積**（透過一致行為循序漸進地增長）、**快速摧毀**（一次欺詐即可災難性地降低分數），以及**個別評估**（每個取用方可套用其自身的聚合函數）。

6.2 可信權威

一個可信權威審查主張，若感滿意則共同簽署——押注其自身的聲譽（圖 7）。這種不對稱是刻意設計的：獲得聲譽緩慢（每次誠實背書遞增 $+\delta$），而喪失它卻是災難性的（每次虛假背書 $-\Delta \gg \delta$；舉例說明，例如 $+2\%$ 對 -70%）。最優策略始終是拒絕可疑的主張。δ 及 Δ 的具體數值是系統參數。

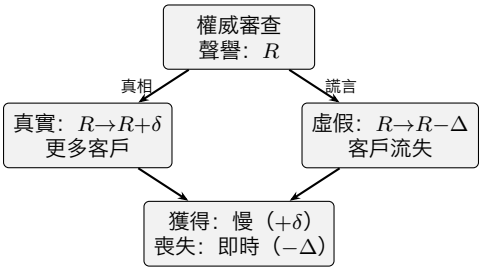

Figure 7: 可信權威——不對稱的誘因。

6.3 多維度聲譽

聲譽並非一個純量，而是一個跨越多個維度的向量：財務可靠性、個人操守、專業能力、公民參與，以及其他（圖 8）。不同的評估供應商會因情境不同而以不同方式投射此向量——放貸者優先考慮財務可靠性，僱主優先考慮專業能力，鄰居優先考慮公民行為。並不存在單一「正確」的聲譽分數；只有視角。

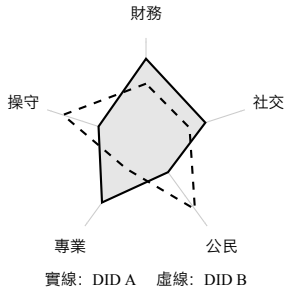

Figure 8: 多維度聲譽向量。不同的 DID 在各維度上呈現不同的輪廓。

6.4 民主化的風險評估

RSN 將系統性風險評估民主化——這是一種目前集中於金融機構、但適用範圍遠超銀行業的能力。工業集團評估供應商可靠性、保險公司為行為風險定價、併購的盡職調查、製造業中的設備壽命估算——但凡交易對手風險需要評估之處，RSN 都提供一個去中心化、市場驅動的替代方案。一個詮釋服務的市場將原始的多維度聲譽資料轉化為可操作的摘要，令交易對手評估對任何參與者而言都能以邊際成本取得。

7 共識與糾紛解決

7.1 去中心化公義模型

RSN 將公義重構為一個雙邊談判問題。永久、可驗證的聲譽損害所帶來的威脅，比受害一方負擔不起的法律程序更為有效的阻嚇。

7.2 湧現式共識

每位參與者維持一個個人的滿意門檻。網絡的總體共識作為數以千計雙邊談判的統計平均值而湧現（圖 9）。一個遠高於共識（野蠻）的回應，發佈起來成本高昂。一個接近共識（相稱）的回應，則成本低廉。共識並非一項規則——它是一個價格信號。

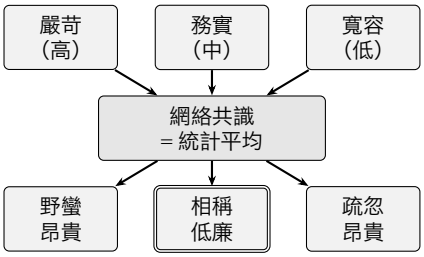

Figure 9: 由個人滿意門檻湧現而生的共識。

7.3 懲罰校準

每位 DID 持有者公開申報對特定不當行為類別的回應。過度嚴苛或過度寬鬆的申報會招致負面的聲譽信號。相稱的申報則不帶摩擦地被接受——一個自我校準的懲罰系統。

共識申報本身亦受偽善偵測所制約：宣示性地承諾一個共識立場，卻行為前後不一，這構成一種比底層偏差更嚴重的聲譽罪過，因為它展示了蓄意的欺騙。

7.4 委託

一個 DID 內的所有活動——除一項例外——均可委託予另一 DID。**對象角色——即主張所關於其行為的 DID——不能委託；它不可轉讓地綁定於主張所指涉的身份持有者。**其他活躍角色——發行者、權威、見證者、驗證者——則可轉移予一個指定的受託 DID，無論是為了驗證、主張提交、共識參與，抑或糾紛解決。委託在任何時候均可撤銷。這使得專業化成為可能（例如專業的驗證服務），同時持有者保留最終的控制權及責任。委託適用於整個系統，並且對可擴展性至關重要。

7.5 從個人道德到湧現式社會契約

每個 DID 的申報政策代表著個人道德的形式化：持有者認為甚麼可接受、他們如何回應特定行為、他們對交易對手有何要求。這些政策並非由系統設計者強加——它們由每位參與者選擇，並以機器可讀的形式表達。

這種形式化促成一個三階段的湧現。首先，個人道德被編碼為**政策**——一套 DID 承諾遵守的申報規則、門檻及回應函數。其次，網絡內所有政策的總和產生出**湧現式倫理**——統計上可觀察的規範，沒有任何單一參與者設計它們，但它們卻由數以千計個人道德立場的互動而生 [9]。第三，這些湧現式倫理，作為透過雙邊價格信號執行的共享行為規範來表達，構成一份**可量度的社會契約**——並非一個無人簽署的理論構造 [10]，而是一套由實際行為支持、可實證觀察的規則。

這一過程與道德基礎理論 [11] 的發現相呼應：人類的道德是直覺的、多元的，並因文化而異。RSN 並不要求道德共識作為輸入；它產生行為共識作為輸出。由此產生的社會契約並非靜態——它隨著參與者加入、離開，並因應網絡回饋調整其政策而演化。與古典社會契約理論不同，這份契約是可撤銷、可觀察，並可在沒有主權者的情況下執行的。

8 經濟模型

前面各節描述了一件工具——RSN 的驗證機制、成本結構及湧現式共識。本節描述一套將此工具應用於財政及治理過渡的方法論。工具是通用的；以下的方法論是一種可能的應用。

8.1 誘因結構

聲譽作為資本，為誠實行為締造直接的誘因。在正常運作中，參與者透過日常交易及履行承諾自然而然地建立聲譽。主要的開支在於負面主張——記錄他人所造成的傷害。這種不對稱激勵有用、可靠的行為：誠實不會帶來額外成本，而造成傷害卻會創造一條有記錄的軌跡，他人會付費將其保存。偽善——申報規則卻不遵守——是最昂貴的失敗模式，因為它展示了蓄意的欺騙。

8.2 並行財政系統

三個組件促成競爭壓力：（1）**簡單稅**——一個沒有任何例外的單一比例稅率，初期略低於現行的有效稅率；（2）**電子開支登記（ESR）**——反轉捷克的 EET 模式，使公民監察國家的開支；（3）**公民導向的稅款分配**——首年由數個百分點起步，並在十年後達到，視乎採用速度而定，由低雙位數百

分比一直到全面遷移至一個公民導向的系統。實際的節奏取決於自由市場對國家服務的替代方案湧現的速度，以及國家配合過渡的意願；時間的函數毋須是線性的，而且沒有理由為採用最終可達之處設定一個上限。

9 遷移路徑

遷移經歷三個階段（圖 10）：**階段 1：覆蓋層**（RSN 作為資訊層，國家是唯一供應者）、**階段 2：競爭**（RSN 提供功能性替代方案，雙系統並用），以及**階段 3：過渡**（RSN 表現超越國家的對等物，自願遷移）。過渡在每一階段都是可逆轉的。

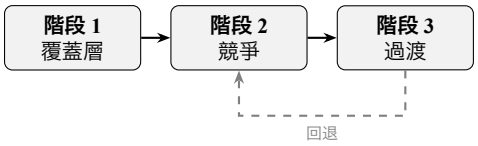

Figure 10: 三階段遷移——在每一階段都可逆轉。

主要的壓力機制是財政上的：當有條件的納稅人達到一個「經濟上具重要性的少數 X」——一個大到足以令針對它的鎮壓造成不容忽視的經濟損害、並使公民抗命成為可信威脅的群體——國家便進入談判。為作說明，我們採用 $X \approx 15\%$ 的 GDP，但實際的門檻取決於具體的經濟體。

10 安全分析

我們考慮五類對手：個別惡意行為者、有組織的群體、企業對手、國家層面的對手，以及協議層面的對手。

女巫攻擊抵抗（Sybil resistance） [12]。建立聲譽需要持續、可驗證的互動。一次成功女巫攻擊的成本隨假身份數量線性增長，隨目標聲譽水平二次方增長。並行運作多個 DID 是可能的，但按設計是昂貴的——每個身份都必須透過真實活動獨立累積聲譽。在自由社會中，這一成本阻嚇隨意的濫用；在獨裁政權下，並行 DID 則成為一種求生工具，使區隔化的抵抗、地下協調及更安全的黑市周旋成為可能。

串謀防禦。封閉群體之間相互背書的模式，透過社交圖譜分析可以偵測。聲譽聚合函數會對來自緊密叢集來源的主張予以折讓。

國家層面的對手。洋蔥路由、中央化基礎設施的缺席，以及驗證者角色分散於參與者群體之中，確保壓制所需的措施與任何收益不成比例。

私隱對問責。化名性——介乎匿名與身份披露之間的一個中間地帶——使 DID 能夠累積有意義的聲譽，而毋須揭露持有者的現實世界身份。

11 私隱考量

RSN 的私隱模型建基於化名性。持有者控制身份的披露：最少（純化名）、選擇性（連結特定憑證），或完全（關聯法律身份）。已發佈的主張是永久的——系統支援情境式更正，但不支援抹除。持有者可放棄一個已被攻陷的 DID 並從頭開始，放棄所累積的聲譽。

12 相關工作

W3C DID Core 規範 [2] 及 Ceramic Network [8] 提供了身份的基礎。EigenTrust [13] 展示了無需中央權威的分散式聲譽聚合。SBT [3] 引入了不可轉讓的社會代幣。RSN 的不同之處在於其強調以經濟成本作為質素過濾器，以及其為激進主義定價的機制。

DAO [4] 及二次方投票 [5] 展示了去中心化治理的可行性。RSN 由雙邊價格談判而非投票中導出共識。

此提案借鑑無政府資本主義理論 [14, 15] 中的私營服務供應，但在兩個關鍵方面有所偏離：它為怨恨及報復性衝動而設計，而非假設理性；並且它提出漸進過渡而非革命。湧現式社會契約的概念在海耶克的自發秩序理論 [16] 中有其先例。

無政府自由市場主義（Anarcho-agorism）。RSN 與自由市場主義的反經濟學（agorist counter-economics） [17] 有相當大的共同之處——尤其是對建立並行機構及自願交換的強調。然而，自由市場主義傾向於假設理性的自利足以作為協調機制，並且往往缺乏一條從現行國家結構出發的具體遷移路徑。RSN 明確地將非理性的行為傾向納入，並提供一種為漸進過渡而設的財政壓力機制。

精英制度（Meritocracy）。RSN 或許代表了對精英制度 [18] 如何能作為一種系統性屬性（而非一個口號）湧現而生的首個功能性描述。傳統的精英制度系統之所以失敗，是因為它們需要一個中央權威來定義及執行「功績」。在 RSN 中，功績由雙邊評估中湧現：那些交付價值的人累積聲譽；沒有委員會裁定誰有功績。

財產作為特權。RSN 從根本上偏離了無政府資本主義及奧地利學派將財產權視為自然且絕對的處理方式。在 RSN 框架中，財產是一種特權——一種社會認可，在極端情況下，當一位參與者從根本上違反共享規範時（背叛、在存亡危機中拒絕捍衛社群、系統性剝削），可被社群收回。這並非國家徵用，而是由雙邊關係網絡收回社會認可。

13 討論與限制

冷啟動。 RSN 的價值取決於網絡效應；早期採用者在參與達到門檻之前面對有限的價值。然而，即使從早期階段起，DID 網絡亦作為一個有用的補充資訊來源。預期早期的聲譽評估及權威評估會隨著日益精密而逐步發展。

反吸血與取用控制。目標 DID 可對來自低聲譽 DID 的查詢施加分級費用及取用限制。這並非二元，而是一個連續的尺度：一個查詢的 DID 聲譽越少，它收到的資訊越少，等待的時間越長，付出的越多。此機制激勵真實的參與，而非被動的取用。

取用的不平等。發佈主張的成本可能會將經濟上處於弱勢的參與者的合理申訴過濾掉。緩解措施：每位參與者同時充當一位潛在的驗證者（或委託此角色），並賺取驗證費用。在足夠長的時間跨度內，一位非激進的參與者應趨近財務中性——驗證收入大致抵銷發佈成本。這是一個統計上的預期，而非保證。

聲譽的不平等。早期採用者累積的優勢，可能為後來者製造障礙。然而，聲譽評估由第三方供應商執行，他們可透過其聚合演算法選擇緩解先行者優勢。以良好聲譽率先起步是一種合理的比較經濟優勢——而這是刻意設計的。

開放問題。最優的成本函數、聚合標準、協議治理，以及與 AI 生成的合成聲譽資料的互動，仍是未來工作的範疇。

本文並不聲稱 RSN 會在所有情況下產生最優的結果。它聲稱的是，RSN 代表一個可行的替代方案——技術上可實行、經濟上能自我維持，並在社會上與人類本來的行為傾向相容。

14 結論

本文提出了聲譽社交網絡，一個使化名、可驗證的聲譽交換成為可能的去中心化協議。昂貴激進主義原則確保欺詐性主張在不施加審查的情況下被定價排除。所提出的遷移路徑使從現行機構出發、漸進且可逆轉的過渡成為可能。此設計將人類本來的天性納入——包括小器、怨恨，以及對報復性滿足的渴求——而非要求意識形態的轉變。其結果並非烏托邦，而是一個公義可及、聲譽靠賺取、而不當行為的代價由造成它的一方承擔的系統。

References

[1] Czech Republic, “Act no. 549/1991 coll., on court fees, as amended,” 1991, fee schedule for court proceedings; cost rules per Act No. 99/1963 Coll. (Code of Civil Procedure) and Decree No. 177/1996 Coll. (Attorney Tariff).

[2] M. Sporny, D. Longley, M. Sabadello, D. Reed, O. Steele, and C. Allen, “Decentralized identifiers (DIDs) v1.0,” W3C Recommendation, Jul. 2022.

[3] E. G. Weyl, P. Ohlhaver, and V. Buterin, “Decentralized society: Finding Web3’s soul,” *SSRN Electronic Journal*, May 2022.

[4] V. Buterin, “DAOs, DACs, DAs and more: An incomplete terminology guide,” Ethereum Blog, May 2014.

[5] V. Buterin, Z. Hitzig, and E. G. Weyl, “A flexible design for funding public goods,” *Management Science*, vol. 65, no. 11, pp. 5171–5187, 2019.

[6] D. Karger, E. Lehman, T. Leighton, R. Panigrahy, M. Levine, and D. Lewin, “Consistent hashing and random trees: Distributed caching protocols for relieving hot spots on the World Wide Web,” in *Proc. 29th ACM STOC*, 1997, pp. 654–663.

[7] D. Buchner, “ION — a scalable DID method based on Bitcoin,” Microsoft, 2021.

[8] Ceramic Network, “Ceramic protocol specification,” GitHub, 2023. [Online]. Available: https://github.com/ceramicnetwork

[9] É. Durkheim, *The Division of Labor in Society*. Free Press, 1893, english translation 1984.

[10] J. Rawls, *A Theory of Justice*. Harvard University Press, 1971.

[11] J. Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Vintage Books, 2012.

[12] J. R. Douceur, “The Sybil attack,” in *Proc. 1st International Workshop on Peer-to-Peer Systems (IPTPS)*, 2002, pp. 251–260.

[13] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, “The EigenTrust algorithm for reputation management in P2P networks,” in *Proc. 12th International Conference on World Wide Web*, 2003, pp. 640–651.

[14] M. N. Rothbard, *For a New Liberty: The Libertarian Manifesto*. Macmillan, 1973.

[15] H.-H. Hoppe, *Democracy: The God That Failed*. Transaction Publishers, 2001.

[16] F. A. Hayek, *Law, Legislation and Liberty*. University of Chicago Press, 1973.

[17] S. E. I. Konkin, *An Agorist Primer*. KoPubCo, 2006.

[18] M. Young, *The Rise of the Meritocracy*. Thames and Hudson, 1958.