去中心化聲譽網路：一個自然社會組織的框架

Pavel Kudrna
Prague, Czechia
初稿 v1 — 2026 年 3 月

Abstract

正義感——即相信錯誤會被糾正、功績會受到回報的信念——是社會最強大的凝聚力量。當中央集權的治理機構因為執行某項權利的成本超過該權利本身的價值而無法提供這種感受時，社會凝聚力便會瓦解。現行的機構結構無法解決一大類糾紛，其失敗方式在系統層面上與中央計劃無法有效配置資源如出一轍：透過資訊不對稱、缺失的回饋迴路，以及決策者與其決策後果之間的脫節。本文提出一種聲譽社交網路（Reputation Social Network，RSN）：一個建立在去中心化識別碼（Decentralized Identifiers，DID）之上、去中心化、不可腐化、抗審查的通訊層，使匿名化參與者能夠針對真實世界的行為發布、驗證並取用聲譽資訊。其核心機制建立在三項設計原則之上：（1）一種非對稱的成本結構，讀取聲譽資料成本低廉，但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定，透過遞增的迭代成本為激進的主張定價；以及（3）一種市場驅動的聲譽權威模型，私人驗證者以其累積的聲譽為其背書之主張的準確性作為抵押。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菁英制（meritocracy），並使得從現有國家管理系統出發的漸進式、可逆的遷移路徑成為可能。

1 緒論

1.1 中央集權式信任的問題

現代治理建立在一個根本的假設之上：中央集權的機構能夠大規模地在陌生人之間居中調解信任。法院裁決糾紛。登記處認證所有權。監理機關執行標準。這些機構是在一個資訊稀缺、通訊緩慢、而將權力委託給中央機構是唯一可行的協調機制的時代所設計的。然而，中央集權式治理失敗的結構性原因與中央計劃相同：資訊不對稱、缺失的回饋迴路，以及決策者與其決策後果之間的脫節。

舉例來說，當機構調解的成本超過糾紛本身的價值時，這個假設便會失效。試想一位租客發現其租賃的公寓缺少法律要求的電氣安全認證——一種對生命構成立即危險的狀況。租客解除合約的法律權利是明確無疑的。然而，透過法院系統執行該權利的成本，卻超過了押金與應賠損害的金額，甚至將潛在的法定賠償計入在內也是如此 [1]。理性的反應就是吞下損失並抽身離開。

這並非病態的極端個案。對於一大類糾紛而言，這才是預設的經歷——這些糾紛中的傷害是真實的，但其金錢利害關係落在機構執行在經濟上變得合理的門檻之下。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動，因為尋求正義的成本落在受害的一方身上。

上述例子取自作者所處的法律環境——捷克的大陸法系。在其他法律傳統中——以判例為基礎的普通法、習慣法、混合體系——正義以不同的方式、不同的程度失效，取決於該司法管轄區的文化與程序特性。讀者很可能會從自身的處境中辨認出類似的例子。在結構上具有普遍性的是這個模式：價值落在經濟理性執行門檻之下的糾紛，最終以受害方吞下損失告終。RSN 並不承諾完美的正義——它僅僅減少當機構執行在經濟上不划算時，惡意行為者所享有的結構性優勢。

1.2 現有解決方案為何不足

去中心化身分（DID）框架 [2] 為自我主權的身分管理提供了密碼學機制，但主要聚焦於身分驗證，並未處理行為聲譽這個更廣泛的問題。

靈魂綁定代幣（Soulbound Tokens，SBTs） [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，且未處理支配資訊寫入的經濟誘因。諸如功績代幣（例如 Liberland）之類的相關概念秉持相似的理念，但仍受制於特定的司法管轄框架。

去中心化自治組織（Decentralized Autonomous Organizations，DAOs） [4] 透過智能合約實現治理，卻複製了金權統治的動態，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。DAOs 還面臨根本的預言機問題（*oracle problem*）：智能合約若無可信的外部來源，便無法可靠地驗證真實世界

的狀態，而這個問題在區塊鏈架構內沒有通用的解決方案。

沒有任何現有系統能同時結合去中心化、抗審查、匿名化、可驗證的寫入成本，以及湧現式的共識。

1.3 貢獻

本文做出以下貢獻:

1. 定義了 **聲譽社交網路（RSN）**，一種擴展 DID 框架以支援雙邊聲譽交換的去中心化協定。
2. 一種基於一致性雜湊 [6] 的 **非確定性驗證者選擇協定**。
3. **昂貴激進主義原則（Expensive Radicalism Principle）**，透過三條疊加的成本管道，依主張與網路共識的距離為其定價。
4. 一種具有非對稱誘因的 **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 **漸進式遷移路徑**。

2 設計原則

RSN 架構受五項不可協商的設計原則所約束。

連續性與演進。系統在一段長度未定的過渡期內與現有機構共存。此過渡必須在每一個階段都可逆。

自願性與責任。參與是自願的，但自願性與責任相互耦合：一位承接某項機構職能控制權的參與者，同時也承擔隨之而來的義務。

多樣性與文化中立。共識湧現自雙邊互動，而非系統設計者強加的預設規則。

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

共識與執行。系統將人類趨向於瑣碎計較、怨恨與報復性滿足的傾向納入為承重的特性。不當行為並非被禁止；它被定價。

3 系統概覽

3.1 聲譽社交網路

RSN 是一個去中心化的對等（peer-to-peer）通訊層，參與者在其中交換關於真實世界行為的可驗證資訊。其定義性屬性為：去中心化且不可審查的基礎設施；透過 DID 進行的匿名化參與；非對稱的成本結構（讀取便宜、寫入昂貴）；密碼學可驗證性；以及 **標準支援**——為透過網路傳播的資訊、為權威所提供的服務（主張組合、背書、見證服務），以及為政策格式（包含評估對手方聲譽的規則，與評估政策不符風險——即宣稱行為與實際行為之間的落差）所提供的機器可讀格式。

3.2 架構組件

RSN 由四個主要組件構成（圖 1）:

1. **DID 層。**參與者身分，含宣告的規則、聲譽中繼資料與網路路由。
2. **主張協定（Claim Protocol）。**用於發布關於真實世界事件之斷言的結構化格式。
3. **驗證網路。**使用一致性雜湊指派驗證者的去中心化協定。
4. **聲譽彙整層。**產生人類可讀之聲譽摘要的服務。

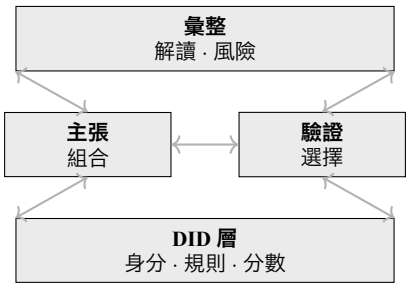

Figure 1: RSN 四層架構。

3.3 參與者角色

一筆驗證交易最多涉及六種不同的 DID 角色（圖 2）：**發起者（Issuer）**（DID_I，建立並提交主張）、**標的（Subject）**（DID_S，主張所針對的 DID——可能與發起者重合）、**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）**（$DID_{W_{1..n}}$，透過挑戰碼隱身監督驗證者誠實度——可選）、**驗證者（Verifier）**（DID_V，以演算法選出以發布主張），以及 **RSN**（已驗證主張得以發布的網路）。任何角色皆可委派給另一個 DID（第 7.4 節）。權威通常會將背書與見證服務綑綁在一起，但這兩項職能在邏輯上是各自獨立的。

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，而是源自四種並存的力量。任何一種都不足以獨立成事；唯有它們的組合才使得腐化的企圖在經濟上變得不理性。

不可預測的目標。每筆主張的驗證者是透過一致性雜湊非確定性地選出的（第 5.3 節）。攻擊者無法預先鎖定某個特定的驗證者行賄，因為驗證者的身分是主張內容與先前迭代的函數，而非發起者所能選擇。

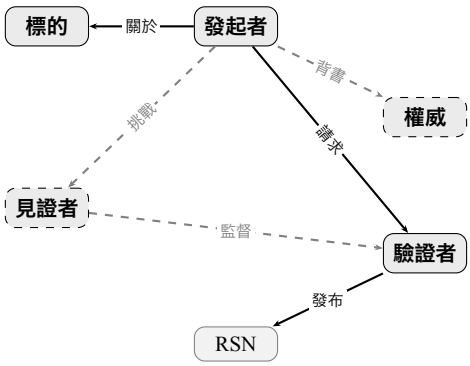

Figure 2: 驗證交易中的 DID 角色。虛線 = 可選（權威、見證者）。

聲譽槓桿——上升緩慢、下墜迅速。聲譽只能透過持續一致的行為漸進地累積。然而，它可以在單一次的詐欺性背書中災難性地喪失（第 6.2 節）。這種非對稱意味著多年累積的聲譽可能被一次不誠實的背書摧毀；最佳策略始終是拒絕可疑的主張。

公開承諾。每一位 DID 持有者都公開宣告其政策——一套他們承諾遵守的機器可讀規則。宣告與實際行為之間的分歧是可偵測的，且被定價為一種比底層違規更嚴重的聲譽罪行（第 7.3 節）。因此，偽善比直接的不誠實代價更高。

網狀攻擊成本非對稱。審查或動搖網路的成本隨網路的規模與密度非線性成長，而容忍它的成本則維持不變。要讓審查划算，一個中央集權的行為者就必須將其施加於整個網路——所需的措施與任何可以想見的收益不成比例（第 10 節）。

4 去中心化身分模型

4.1 具有特殊屬性的 DID

RSN 擴展了 W3C DID Core 規範 [2]，加入：**宣告的規則**（機器可讀的行為承諾）、**聲譽中繼資料**（彙整後的行為分數），以及 **網路路由**（可變的洋蔥網路閘道，與穩定的環上位置相分離）。

4.2 主張生命週期

一筆主張經歷六個階段（圖 3）：組合、可選的權威背書、透過一致性雜湊進行的驗證者選擇、驗證決策（接受或拒絕並進行迭代）、發布，以及對所有各方的聲譽更新。

4.3 與 W3C DID Core 的關係

RSN 的身分模型是 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的 DID 方法規範中所定義的 DID 文件屬性，與 ION [7] 和 Ceramic [8] 等現有基礎設施相容。

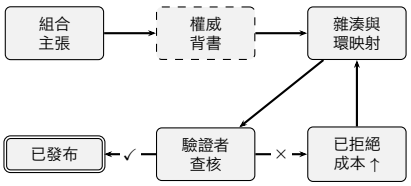

Figure 3: 含迭代迴路的主張生命週期。

5 資訊驗證與傳播

5.1 資訊寫入成本

RSN 的核心經濟原則是讀取與寫入之間的非對稱。對於身處 DID 網路之中、且擁有與其聲譽相稱之取用權限的參與者而言，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（見反吸血機制）。寫入——理解為將聲譽持久地實體化進網路，而非一次性的技術行為——需要跨三個維度的可驗證累積投入：**時間**（一生中花在一致行為、履行承諾與經營關係上的歲月）、**精力**（投入於誠實往來、關照夥伴關係與長期可信賴度的付出），以及 **金錢**（對自身名譽的投資——專業發展、自身作品的品質、對所致傷害的賠償）。聲譽無法即時買到——它是終身累積的成果，系統僅僅使其變得可見，並可跨情境轉移。協定的一次性技術成本（密碼學簽章、驗證者費用）相較於這種底層投資微不足道。

5.2 社交圖譜與聯絡深度

RSN 本質上是一個社交網路：每個 DID 加入允許此連結的聯絡人（其他 DID）。這些聯絡人又有他們自己的聯絡人，以此類推。演算法驗證者搜尋在可設定的關聯聯絡人深度 h 之內運作（例如 $h = 3$：自己的直接聯絡人、他們的聯絡人，以及聯絡人的聯絡人）。此架構不需要單一的全球區塊鏈——它自然地偏好社群／叢集的湧現，且與其他社群有所重疊。每一位 DID 參與者都必須有一份已發布的 **政策**——一套支配如何評估傳入請求的機器可讀規則。政策違反會作為負面產物記錄於網路中。

5.3 演算法驗證者選擇

驗證協定採用一致性雜湊 [6]（圖 4）。驗證是一項付費服務：驗證者收費運作，形成一個市場，其中競爭驅動了定價、速度與可靠性。

步驟 1。主張文件與前一次迭代的雜湊結果（或第一次迭代時為 $\varnothing$）串接後進行雜湊：

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

演算法必須納入所有先前請求與回應的完整路徑——而不僅僅是前一次迭代的雜湊。每一位驗證者的完整回應（接受、拒絕、報價、理由）都被附加到不斷增長的文件中，並在每一步都可透過密碼學簽章加以驗證。

步驟 2。雜湊被映射到一致性雜湊環上的 N 個位置，均勻分布（例如當 $N = 4$: $+0°, +90°, +180°, +270°$）。從每個位置出發，順時針搜尋選出最近的 DID 作為驗證者候選（圖 5）。N 是一個可變參數，可能取決於當前活躍 DID 的百分比、一天中的時段，或網路歷史上的回應能力。

步驟 3。驗證者接受（發布，以聲譽作為抵押）或拒絕（帶著拒絕雜湊回到步驟 1）。當某個節點雖宣告為線上狀態卻不予回應時，下一次請求必須納入來自所有 N 個先前驗證者候選的全部回應。

反吸血。標的 DID 可對來自聲譽甚少之新 DID 的查詢設定費用或取用限制。這種漸進式（而非二元的）機制迫使新加入者透過真實的活動建立聲譽，然後才能自由取用他人的資料。

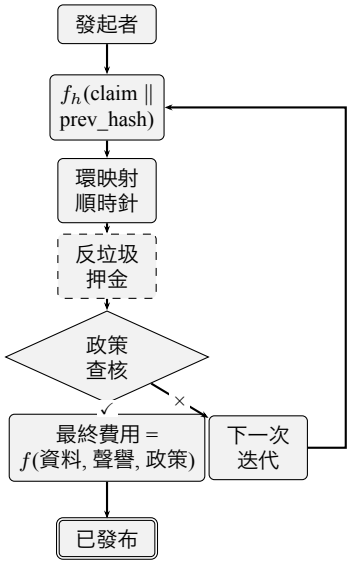

Figure 4: 驗證者選擇協定。反垃圾押金為可選且可能可退還。最終費用僅在接受時支付。

每一次拒絕都會將驗證者的完整回應（包含理由與條件）附加到主張文件上，擴充其內容。從擴充後的文件中，會非確定性地計算出一個新的雜湊，映射到不同的環上位置並選出不同的驗證者。完整路徑的記錄是強制性的——發起者無法預測或影響下一個驗證者。若某位驗證者儘管有相容的宣告政策卻無視請求，見證者（若在場）會記錄此事，發起者可在最終發布時附上見證證據。

5.4 見證協定

見證者角色透過密碼學挑戰碼協定，為驗證者的誠實度提供隱身的問責：

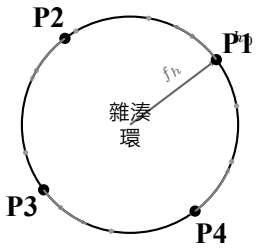

Figure 5: 多點選擇（N=4）。雜湊 h_0 設定偏移量；4 個點從 h_0 均勻分布。

步驟 W1。請求者簽署驗證請求並將其發送給 N_w 位見證者——來自請求者社交網路的聯絡人，或收費運作的專門見證服務提供者。

步驟 W2。每位見證者 w_i 簽署整份已簽署的請求，保留原始簽章 σ_i，並計算出一個挑戰碼 $c_i = h(\sigma_i)$。此挑戰碼被附加到請求上。

步驟 W3。如今帶有 N_w 個挑戰碼的請求被發送給驗證者。驗證者觀察到這些碼，但無法判定誰是見證者，或這些碼是否對應到真實的簽章。

步驟 W4（誠實的驗證者）。若驗證者依其宣告的政策回應，挑戰碼便維持不透明。見證者永不揭露。不發布任何額外資料。

步驟 W5（政策違反）。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章 σ_i。這些可作為佐證資料發布：任何人都能驗證 $c_i = h(\sigma_i)$，證明見證者當時在場。請求者可以獨立發布——驗證者早已在其回應中簽署了這些挑戰碼。

虛張聲勢的特性。請求者可以用隨機值替換部分或全部的挑戰碼。驗證者無法區分真正的碼（由高聲譽的權威作為後盾）與雜訊。這種不確定性正是執行機制：每一個請求都帶有一個風險，即某位受人敬重的權威正隱身觀看。維持這種壓力的成本近乎於零（生成隨機數），而驗證者若不誠實的潛在代價則是災難性的。即使在實際上沒有見證者的情況下，誠實的行為也受到激勵。

權威作為見證者。一位為主張背書的權威可以綑綁見證服務：「我為你的主張背書，並在驗證期間作為隱身見證者。」這是一種自然的商業模式——權威本就已掌握脈絡。然而，這兩種角色在邏輯上是各自獨立的。

5.5 昂貴激進主義原則

三條成本管道同時疊加（圖 6）：

管道 1：迭代成本。每一次拒絕都迫使一次消耗時間與資源的新迭代。線性成長。

管道 2：文件膨脹。每一次迭代都將驗證者的完

整回應加入主張文件。成長是線性的，但帶有一個基底偏移：初始的酬載（主張內容、權威背書、密碼學簽章）本就已具有不小的大小 B_0。k 次迭代後的總大小：$S(k) = B_0 + k \cdot \Delta$，其中 Δ 為平均回應大小。

管道 3：驗證者風險溢價。風險是主觀的。驗證者評估請求方 DID 的宣告政策是否與驗證者自身的商業、社會或政治利益相衝突。此溢價可能隨與驗證者「理想」的感知距離指數式攀升：$P(d) = \alpha \cdot e^{\beta d}$，其中 d 為驗證者的主觀距離度量。超過某個個人的禁區邊界後，驗證者可能完全拒絕。

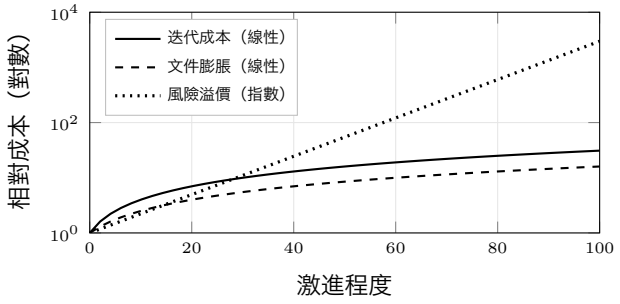

Figure 6: 三條管道上的成本攀升。風險溢價在高激進程度時佔主導地位。

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: 各類主張的成本比較。

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

6 聲譽與風險評估

6.1 聲譽累積模型

聲譽展現三種屬性：**緩慢累積**（透過一致行為漸進成長）、**迅速摧毀**（單一次詐欺就能災難性地降低分數），以及 **個別評估**（每一個取用方都可套用自己的彙整函數）。

6.2 可信權威

一位可信權威審查主張，若感到滿意便共同簽署——以其自身的聲譽作為抵押（圖 7）。這種非對稱是刻意設計的：獲得聲譽是緩慢的（每次誠實背書漸進 $+\delta$），而喪失聲譽則是災難性的（每次虛假背書 $-\Delta \gg \delta$；舉例說明，例如 $+2\%$ 對比 -70%）。最佳策略始終是拒絕可疑的主張。δ 與 Δ 的具體數值為系統參數。

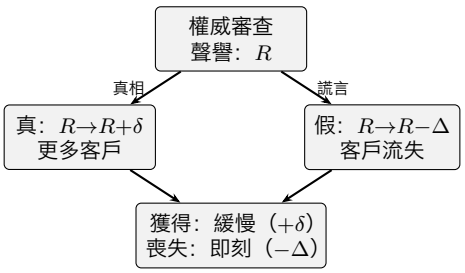

Figure 7: 可信權威——非對稱誘因。

6.3 多維度聲譽

聲譽並非一個純量，而是一個橫跨多個維度的向量：財務可靠性、個人正直、專業能力、公民參與等等（圖 8）。不同的評估提供者依情境以不同方式投影這個向量——放款者優先考量財務可靠性，雇主優先考量專業能力，鄰居優先考量公民行為。並不存在單一「正確」的聲譽分數；只有各種觀點。

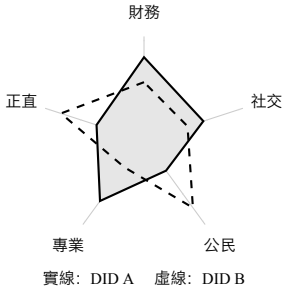

Figure 8: 多維度聲譽向量。不同的 DID 在各維度上展現不同的輪廓。

6.4 民主化的風險評估

RSN 使系統性風險評估民主化——這種能力目前集中於金融機構之手，但其適用範圍遠遠超出銀行業。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——凡是需要評估對手方風險之處，RSN 都提供了一種去中心化、市場驅動的替代方案。一個解讀服務的市場將原始的多維度聲譽資料轉譯為可操作的摘要，使任何參與者都能以邊際成本進行對手方評估。

7 共識與糾紛解決

7.1 去中心化正義模型

RSN 將正義重新框定為一個雙邊協商問題。永久、可驗證的聲譽損害這一威脅，比受害方無力負擔的法律程序更能有效嚇阻。

7.2 湧現式共識

每一位參與者都維持一個個人的滿意門檻。網路的整體共識湧現為數以千計的雙邊協商的統計平均值（圖 9）。遠高於共識的回應（野蠻的）發布起來昂貴。接近共識的回應（相稱的）便宜。共識不是一條規則——它是一個價格信號。

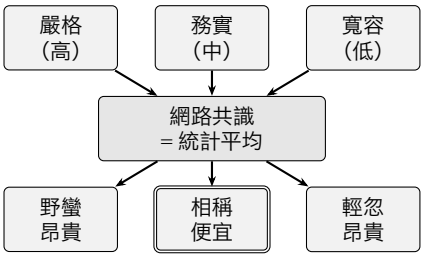

Figure 9: 由個人滿意門檻湧現的共識。

7.3 懲罰校準

每一位 DID 持有者都公開宣告其對特定不當行為類別的回應。過於嚴苛或過於寬鬆的宣告會招致負面的聲譽信號。相稱的宣告則毫無摩擦地被接受——一套自我校準的懲罰系統。

共識宣告本身也受到偽善偵測的約束：宣告性地承諾某個共識立場，卻表現得前後不一，構成一種比底層偏差更嚴重的聲譽罪行，因為它展現了蓄意的欺騙。

7.4 委派

一個 DID 之內的所有活動——除了一個例外——皆可委派給另一個 DID。**標的角色——即主張所針對其行為的 DID——不可委派；它不可轉讓地綁定於主張所指涉的身分持有者。**其他的主動角色——發起者、權威、見證者、驗證者——皆可移轉給一個指定的受託 DID，無論是為了驗證、主張提交、共識參與，或糾紛解決。委派可隨時撤銷。這使得專業化（例如專業的驗證服務）成為可能，同時持有者保有最終的控制權與責任。委派適用於整個系統，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

每一個 DID 所宣告的政策代表著個人道德的一種形式化：持有者認為何者可接受、他們如何回應特定行為、他們對對手方有何要求。這些政策並非由系統設計者強加——它們由每一位參與者所選擇，並以機器可讀的形式表達。

這種形式化促成了一種三階段的湧現。首先，個人道德被編碼為 **政策**——一套 DID 承諾遵守的宣告規則、門檻與回應函數。其次，網路上所有政策的總和產生了 **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了它，而它卻源自數以千計個人道德立場的互動 [9]。第三，這些湧現式倫理，表現為透過雙邊價格信號執行的共享行為規範，構成了一份 **可衡量的社會契約**——不是一個沒有人簽署的理論建構 [10]，而是一套經驗上可觀察、由實際行為作為後盾的規則。

這個過程呼應了道德基礎理論 [11] 的發現：人類的道德是直覺的、多元的，且因文化而異。RSN 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的——它隨著參與者的加入、離開，以及依網路回饋調整其政策而演化。與古典的社會契約理論不同，這份契約是可撤銷、可觀察，且無需一位主權者即可執行的。

8 經濟模型

前面各節描述了一項工具——RSN 的驗證機制、成本結構與湧現式共識。本節描述一套將此工具應用於財政與治理轉型的方法論。工具是通用的；下述方法論則是其一種可能的應用。

8.1 誘因結構

聲譽作為資本，為誠實行為創造了直接的誘因。在正常運作中，參與者透過日常交易與履行的承諾自然地建立聲譽。主要的支出用於 負面主張——記錄他人所造成的傷害。這種非對稱激勵了有用、可靠的行為：誠實不花費額外成本，而傷害人則會製造出一條有據可查的軌跡，他人會付費將其保存下來。偽善——宣告規則卻不遵守——是代價最高的失敗模式，因為它展現了蓄意的欺騙。

8.2 並行財政系統

三個組件促成了競爭壓力：（1）**簡單稅**——一種毫無例外的單一比例稅率，起初略低於現行的有效稅率；（2）**電子支出登記（Electronic Spending Registration，ESR）**——反轉捷克的 EET 模式，使公民得以監督國家支出；（3）**公民導向的稅收分配**——第一年從百分之幾起步，十年後達到從低雙位數百分比、一直到完全遷移至公民導向系統之間的任何程度，取決於採用的速率。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，以及國家配合此過渡的意願；時間的函數不必是線性的，也沒有理由為採用最終可能抵達之處設定一個上限。

9 遷移路徑

遷移分三個階段進行（圖 10）：**階段 1：疊加**（RSN 作為資訊層，國家為唯一提供者）、**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：轉型**（RSN 表現勝過國家的對等物，自願遷移）。此過渡在每一個階段都可逆。

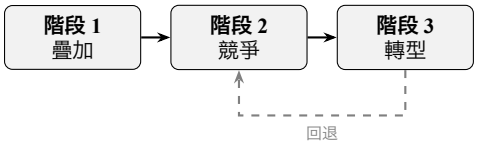

Figure 10: 三階段遷移——在每一個階段都可逆。

主要的壓力機制是財政性的：當有條件的納稅人達到一個「經濟上顯著的少數 X」——一個大到足以使針對它的鎮壓造成不小的經濟損害、且公民不服從成為可信威脅的群體——時，國家便進入協商。為了說明，我們採用 $X \approx 15\%$ 的 GDP，但實際的門檻取決於特定的經濟體。

10 安全分析

我們考量五類對手：個別惡意行為者、有組織的群體、企業對手、國家級對手，以及協定級對手。

女巫（Sybil）抵抗力 [12]。建立聲譽需要持續、可驗證的互動。一次成功的女巫攻擊的成本隨假身分數量線性成長，並隨目標聲譽水準二次方成長。並行操作多個 DID 是可能的，但依設計而言代價高昂——每一個身分都必須透過真實的活動獨立地累積聲譽。在自由社會中，這種成本嚇阻了隨意的濫用；在獨裁政權下，並行的 DID 則成為一種求生工具，使得區隔化的抵抗、地下協調，以及更安全的黑市穿行成為可能。

共謀防禦。封閉群體之間相互背書的模式可透過社交圖譜分析加以偵測。聲譽彙整函數會對來自緊密叢集之來源的主張予以折抵。

國家級對手。洋蔥路由、中央基礎設施的缺席，以及驗證者角色橫跨參與者群體的分布，確保了壓制需要與任何收益不成比例的措施。

隱私對比問責。匿名化——介於匿名與身分揭露之間的一個中間地帶——使得 DID 得以累積有意義的聲譽，而無需揭露持有者的真實世界身分。

11 隱私考量

RSN 的隱私模型建立在匿名化之上。持有者控制身分的揭露：最小（純粹匿名代號）、選擇性（連結特定憑證），或完整（關聯法律身分）。已發布的主張是永久的——系統支援情境式的更正，但不支援抹除。持有者可以放棄一個已遭破解的 DID 並重新開始，代價是放棄累積的聲譽。

12 相關工作

W3C DID Core 規範 [2] 與 Ceramic Network [8] 提供了身分的基礎。EigenTrust [13] 展示了無須中央權威的分散式聲譽彙整。SBTs [3] 引入了不可轉讓的社交代幣。RSN 的不同之處在於它強調以經濟成本作為品質過濾器，以及其為激進主義定價的機制。

DAOs [4] 與二次方投票 [5] 展示了去中心化治理的可行性。RSN 從雙邊價格協商而非投票中導出共識。

本提案借鑑了無政府資本主義理論 [14, 15] 中關於私人服務提供的部分，但在兩個關鍵之處有所偏離：它為怨恨與報復性衝動而設計，而非假設理性；並且它提出漸進式過渡而非革命。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 與自由市場反經濟學（agorist counter-economics）[17] 有著重大的共通之處——尤其是對建立並行機構與自願交換的強調。然而，自由市場主義傾向於假設理性的自利是一種充分的協調機制，且往往缺乏一條從現有國家結構出發的具體遷移路徑。RSN 明確地納入了非理性的行為傾向，並提供了一套用於漸進過渡的財政壓力機制。

菁英制。 RSN 或許代表了對菁英制 [18] 如何能作為一種系統性屬性而非一句口號湧現的首次功能性描述。傳統的菁英制系統之所以失敗，是因為它們需要一個中央權威來定義並執行「功績」。在 RSN 中，功績湧現自雙邊評估：那些交付價值的人累積聲譽；沒有任何委員會決定誰具有功績。

財產作為特權。 RSN 從根本上偏離了無政府資本主義與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，當某位參與者從根本上違反共享規範（背叛、在存亡危機中拒絕捍衛社群、系統性剝削）時，社群可以將其撤回。這並非國家徵收，而是雙邊關係網路對社會認可的撤回。

13 討論與限制

冷啟動。 RSN 的價值取決於網路效應；早期採用者在參與度達到某個門檻之前面臨有限的價值。然而，即便從早期階段起，DID 網路也可作為一種有用的補充資訊來源。早期的聲譽評估與權威評估預期會隨著日益成熟而逐步發展。

反吸血與取用控制。標的 DID 可對來自低聲譽 DID 的查詢施加漸進式的費用與取用限制。這並非二元的，而是一個連續的尺度：查詢方 DID 的聲譽越少，它收到的資訊越少、等待越久、付出越多。此機制激勵了真實的參與，而非被動的消費。

取用的不平等。發布主張的成本可能會將經濟弱勢參與者的正當申訴過濾掉。緩解方式：每一位參與者同時作為潛在的驗證者（或委派此角色）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於財務中性——驗證收入大致抵銷發布成本。這是一種統計上的預期，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，聲譽評估由第三方提供者執行，他們可以選擇透過其彙整演算法緩解先行者優勢。率先擁有良好聲譽是一種正當的比較性經濟優勢——而這正是刻意的設計。

開放問題。最佳成本函數、彙整標準、協定治理，以及與 AI 生成的合成聲譽資料的互動，仍是未來工作的領域。

本文並不主張 RSN 會在所有情況下產生最佳結果。它主張 RSN 代表一種 可行的替代方案——在技術上可實現、在經濟上可自我維持，並在社會層面上與如其實際樣貌的人類行為傾向相容。

14 結論

本文提出了聲譽社交網路，一種使匿名化、可驗證的聲譽交換成為可能的去中心化協定。昂貴激進主義原則確保詐欺性主張在不加審查的情況下因價格過高而被排除。所提出的遷移路徑使得從現有機構出發的漸進、可逆的過渡成為可能。此設計將如其實際樣貌的人性——包含瑣碎計較、怨恨，以及對報復性滿足的渴望——納入其中，而非要求意識形態的轉變。其結果不是烏托邦，而是一個正義可及、聲譽靠賺得、且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.