Human-Centered Decision Support for Crypto Airdrop Governance: Interpretable Behavioral Modeling of Strategic Hunters

Lindsay Zastrow, Cheri Allen, Leah Ingold, Adeline Wen, Yan Bai, and Wei Cai

***Abstract*— Cryptocurrency airdrops (free token giveaways to attract early users) are routinely exploited by strategic actors who create many wallets to claim rewards repeatedly. For operators, the challenge is not only identifying suspicious accounts, but configuring rules they can understand, audit, justify, and adapt across campaigns. Using 2.5 million address-project records from 16 Ethereum projects, we analyze every possible combination of five human-understandable behavioral indicators and quantify each indicator's contribution under an internal-label utility. Three findings characterize this rule-design space: (1) under the selected internal-label utility, fund-flow indicators receive the largest Shapley contribution shares; (2) using HasciDB's five-indicator OR label as the reference, the HF+RF+MA subset retains 78% of positive address-project records; and (3) externally listed address groups exhibit different indicator-trigger profiles in their available HasciDB histories. The external comparison measures flagged-list agreement rather than independent validation of detection accuracy. The result is a compact, auditable summary for examining and documenting candidate screening configurations, not a validated universal defense.**

I. INTRODUCTION

Token airdrops are a common way for blockchain projects to attract users: project teams distribute free tokens to early participants [1]. The mechanism is attractive precisely because it is open and easy to join. The same openness also makes it easy to exploit. One person can register hundreds or thousands of wallet addresses, appear to be many different people, and claim rewards repeatedly. This is a Sybil attack [2]. For project operators, the real issue is not just abuse, but that decisions to exclude suspicious accounts are public, consequential, and often contested.

Thus, airdrop abuse is as much a human-centered decision problem as a technical detection problem. Operators must choose what to monitor, how strict a policy should be, how to explain exclusions, and how to adapt rules across campaigns. Accurate but opaque detectors are difficult to audit and defend, whereas simple but poorly tuned rules may be too weak. Operators need compact rule sets they can understand, justify, and adapt as tactics change.

Existing research has used machine learning to identify suspicious wallets [3], [4] and graph-based forensics to trace hidden coordination through on-chain fund flow [5], [6], [7], [8]. These methods are valuable for ranking or detecting suspicious accounts, but they usually foreground predictive performance over operator-facing transparency. Risk-scoring and identity systems can label an account as trusted or risky, yet often do not tell a reviewer what concrete behavior led to that judgment. For governance-facing deployment, the missing layer is an interpretable behavioral representation that supports human rule selection.

This paper addresses that gap by treating airdrop hunters as strategic actors who respond to public rules and incentives. The adaptive-agent lens is behavioral, not a formal cognitive model: we do not observe internal states, and this cross-sectional analysis does not test longitudinal adaptation. We describe suspicious activity with five human-understandable indicators from prior work [1]; rather than black-box features, they state which explicit rule conditions a record triggers. Using HasciDB (`https://hascidb.org`), a dataset with over 2.5 million address-project records from 16 Ethereum projects, of which 38.0% are positive under the internal five-indicator OR label, we test all possible indicator combinations, compute exact Shapley values for indicator contribution, compare rules of varying strictness, and examine agreement with three external flagged-address lists. Our goal is not only to measure internal rule retention, but to show how the evidence can support human-selectable, auditable rule design. To make the study objectives explicit in the introduction, we pose three research questions: RQ1, which indicators contribute most under the selected internal-label utility; RQ2, how internal-label retention changes with rule-set size or simultaneous-trigger requirements; and RQ3, how external-list agreement varies and what this implies for context-specific configuration. Because the internal HasciDB label is constructed from the same five indicators, these are policy-design questions rather than claims about independent ground-truth detection accuracy.

The main contributions of this paper are:

- We present an interpretable behavioral representation of strategic airdrop abuse using five operator-readable indicators, framing the problem as human-centered rule design rather than only automated detection.
- We compute exact Shapley values over all 31 non-empty indicator combinations under a defined internal-label utility, transparently attributing each rule's contribution to reproducing the HasciDB OR label.
- We show that most internally flagged address-project records trigger exactly one indicator and quantify how retention declines when multiple simultaneous activations are required.
- We characterize agreement between Gitcoin, Hop, and LayerZero flagged-address lists and indicator activations in the listed addresses' available HasciDB histories.

*This work was partially supported by National Science Foundation (NSF) with award number 2334196 and 1921576.

All authors are with the School of Engineering and Technology, University of Washington, Tacoma, WA, 98402 USA. Emails: {lzastrow, ca27, lingold, ywen8, yanb, weicaics}@uw.edu.

© 2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. Accepted manuscript of: Lindsay Zastrow, Cheri Allen, Leah Ingold, Adeline Wen, Yan Bai, Wei Cai, "Human-Centered Decision Support for Crypto Airdrop Governance: Interpretable Behavioral Modeling of Strategic Hunters," in 2026 IEEE International Conference on Systems, Man, and Cybernetics (SMC'26), Bellevue, WA, USA, October 4-7, 2026.

II. RELATED WORK

This study connects four research directions that matter for operator-facing airdrop governance.

Interpretable screening and decision support. In practice, operators need more than a binary risk label: they need a screening basis they can inspect, explain, and defend. Risk-scoring platforms such as Nansen and TrustaLabs, together with identity-verification systems like Human Passport and World ID, support trust assessment in crypto ecosystems [9], [10], [11], [12]. However, these systems typically answer whether an account appears risky or legitimate, not which behavioral signals should drive transparent policy. Our work contributes at this decision-support layer by turning suspicious activity into operator-readable behavioral rules.

Classical Sybil defense. Early Sybil-defense work treated the problem as identity fabrication and proposed using social-network structure to bound the influence of fake identities (e.g., SybilGuard, SybilLimit) [13], [14], [15]. Those methods were not designed for public blockchain environments, where wallet creation is cheap and strategic actors can adapt directly to visible rules. We inherit the concern with fake identities, but focus on how human operators can configure understandable defenses in a low-friction adversarial setting.

Blockchain-native detection and forensics. Researchers analyze transaction graphs and fund-flow structure to uncover hidden coordination [5], [6], [7], [8]. More recent airdrop-focused work uses graph learning and classifiers to improve detection performance [3], [4]. These methods are effective at ranking suspicious wallets, but they usually optimize for predictive performance rather than transparent rule analysis. Our work complements them by measuring internal-label retention, overlap, and marginal utility across small interpretable signal sets, while leaving policy selection to campaign-relevant validation.

Strategic abuse under rule-based incentives. Recent work shows that airdrop-hunter behavior is shaped by campaign rules and reward design, with different campaigns eliciting different strategies [1], [2]. We extend that direction by treating behavioral indicators as governance-facing decision variables: we quantify each indicator's internal-label contribution, compare rule strictness, and analyze project-level trigger distributions, while leaving policy selection to campaign-specific validation.

III. DATASET AND PROBLEM SETTING

We use the HasciDB dataset. It contains 2,539,857 records from 16 Ethereum projects. Since one address can participate in multiple projects (producing multiple records), after deduplication there are 1,807,415 unique addresses. The internal five-indicator OR-label positive rate is 38.0% (964,034 address-project records). Observed per-project internal flag rates range from 5.5% to 75.4%; these descriptive rates are not strictly comparable across all projects because candidate-population construction and chain coverage differ.

The Sybil labels come from five behavioral indicators established in prior work [1]. That study surveyed blockchain security practitioners, collected candidate indicators of hunter behavior, and refined them through multiple rounds of structured voting (Delphi method) until the panel agreed on five indicators intended to characterize behaviors treated as suspicious within the HasciDB rule ontology. Table I lists each indicator's meaning, threshold, and trigger rate.

TABLE I
THE FIVE BEHAVIORAL INDICATORS IN HASCIDB

Code	Meaning	Threshold	Fire Rate
BT	Automated scripted trading	≥ 5 txns	10.7%
BW	Batch new-wallet activation	≥ 10 wallets	3.2%
HF	High-frequency trading arbitrage	≥ 0.80	8.2%
RF	Rapid fund consolidation	≥ 0.50	14.2%
MA	Multi-address circular arbitrage	≥ 5 loops	11.3%

HasciDB groups these into two categories: operations (BT, BW, HF) and fund-flow (RF, MA). If any indicator in either category fires for an address-project record, that record receives a positive value under HasciDB's internal OR label:

```
ops_flag = (BT≥5) OR (BW≥10) OR (HF≥0.80)
fund_flag = (RF≥0.50) OR (MA≥5)
is_sybil = ops_flag OR fund_flag
```

This grouping is purely organizational; the net effect is "flag if any of the five indicators fires." The pipeline also excludes about 174,000 known exchange and infrastructure addresses (from Dune Analytics public labels) to avoid flagging centralized platforms. HasciDB additionally contains public token allocation data (how many tokens each address claimed) for projects that disclosed this information.

Important note: because the `is_sybil` label is computed from these same five indicators, every candidate rule subset is nested within the parent OR definition. Precision relative to that internal label is 100% by construction; this is set-containment, not empirical precision against independently verified identities. We report internal-label retention and rule overlap rather than independent detection accuracy.

Our experimental design is therefore organized around three operator questions:

Question 1 (What should an operator monitor?) We test all 31 indicator combinations and find the highest-retention subset at each size. This identifies candidate *rule sets*, but not yet the marginal value of each *individual* signal.

Question 2 (How simple or strict can the policy be?) Shapley values answer the individual-contribution question by averaging over all possible combinations. We also measure how internal-label retention changes when multiple indicators must fire simultaneously. Together, these analyses provide descriptive inputs for rule review; they do not validate a deployment choice.

Question 3 (How does external flagged-list agreement vary?) We examine whether addresses on the Gitcoin, Hop, and LayerZero lists have triggered the selected indicators in their available HasciDB histories.

We also classify the 16 projects by sector (Table II) to organize the project-level trigger distributions reported later;

sector differences are descriptive and do not establish how real actors behave or why the distributions differ.

TABLE II
PROJECT SECTOR CLASSIFICATION

Sector	Projects	Count
Trading	uniswap, 1inch, dydx, paraswap	4
NFT	looksrare, x2y2, blur_s1, blur_s2, apecoin, pengu	6
Staking	eigenlayer, etherfi, badger	3
Infra/Gov	ens, gitcoin, ampleforth	3

IV. METHODOLOGY

A. Testing All Indicator Combinations

The first operator question is what to monitor. Five indicators produce 31 non-empty combinations (each indicator is either included or not, giving $2^5 = 32$ possibilities, minus the empty set where nothing is used). For each combination, we flag an address-project record if any included indicator fires and calculate the fraction of records positive under the full five-indicator OR label that the subset retains. We call this quantity internal-label retention. Because every subset output is contained within the parent OR label by definition, this analysis does not estimate empirical correctness against independent ground truth.

B. Shapley Values: Measuring Each Indicator's Contribution

The second operator question is how marginal internal utility is distributed across signals in a candidate rule set. Shapley values come from game theory and were originally designed to fairly divide a team's payoff. The idea is: if a player joins different team compositions, how much does performance improve on average? That average improvement is the player's contribution.

In our setting, the "players" are the five indicators and the value function is $v(S) = 2r/(1 + r)$, where r is a subset's internal-label retention. This is the standard F1 formula under the parent-label containment property, so it is a monotonic transformation of retention rather than an independently estimated detection score. We use it to assign diminishing marginal weight to later retention gains; it does not directly encode rule-set size, implementation cost, or governance cost. Let $P = \{BT, BW, HF, RF, MA\}$ denote the full indicator set. For one focal indicator i, the Shapley value is:

$$\phi_i = \sum_{S \subseteq P \setminus \{i\}} \frac{|S|!(|P| - |S| - 1)!}{|P|!} \left[v(S \cup \{i\}) - v(S) \right], \tag{1}$$

where S is any subset that does not already contain i, and $v(S)$ is the F1 score of combination S. The term $v(S \cup \{i\}) - v(S)$ is the marginal gain from adding indicator i to that particular combination. Shapley values average this marginal gain over all possible subsets S, with the combinatorial weight ensuring each possible ordering of indicators is treated fairly. Since there are only 32 possible subsets (including the empty set), we can compute this exactly without approximation.

For example, when computing HF's contribution, $i = HF$ and S ranges over subsets of $\{BT, BW, RF, MA\}$, such as $\varnothing$, $\{BT\}$, or $\{RF, MA\}$. The five Shapley values sum to the full-set F1. We divide each by the total and multiply by 100% to get each indicator's percentage share. A higher share means that the indicator provides a larger average marginal increment under this selected internal-label utility; it is not independently validated detection importance.

Why not just pick the highest-retention combination? Shapley values show how marginal retention and overlap are distributed across rules under the selected utility.

C. Voting Rules: How Many Indicators Must Fire?

After finding the highest-retention combination at each size, we ask whether one, two, or three selected indicators must fire simultaneously. This choice determines how many explicit rule conditions are required before a record is flagged for review; it is not, by itself, a validated standard for exclusion.

- OR-1: flag if at least one indicator fires.
- OR-2: flag if at least two indicators fire simultaneously.
- OR-3: flag if at least three indicators fire simultaneously.

Among the 686,893 unique addresses with at least one OR-positive HasciDB record, we aggregate project-level observations into one row per address using ever-triggered binary indicators and compute pairwise Pearson correlations. Because inclusion in this address set is itself defined by the union of the same indicators, these are conditional descriptive associations and the selection can mechanically shift correlations in the negative direction.

The voting analysis measures how much of the parent internal label remains as the simultaneous-trigger requirement increases. It does not identify unique actors or the number of real-world techniques they use.

D. Agreement with External Flagged-Address Lists

The third operator question concerns the evidence boundary for reusing a rule set. We compare against flagged-address lists published for Gitcoin, a grants-allocation platform [16]; Hop, a cross-chain bridge protocol [17]; and LayerZero, an omnichain interoperability protocol [18]. Each was independently compiled by its respective project team.

Method: for addresses appearing in both an external list and HasciDB, we determine whether they triggered selected indicators in their available HasciDB project histories. We do not recompute the indicators within the external campaigns, and the external data provide no independently verified negative class. The resulting percentages measure flagged-list agreement rather than out-of-sample detector validation.

V. RESULTS

A. Fund-Flow Indicators Receive the Largest Internal-Utility Shares

Fig. 1 shows each indicator's Shapley contribution share: RF (fund consolidation) 29.0%, MA (circular arbitrage)

23.3%, BT (scripted trading) 22.4%, HF (high-frequency activity) 18.3%, BW (batch wallet creation) 7.0%. Under the selected internal-label utility, RF and MA receive the largest shares. This ranking describes marginal contribution to reproducing the HasciDB OR label, not independently verified accuracy for identifying real-world hunters.

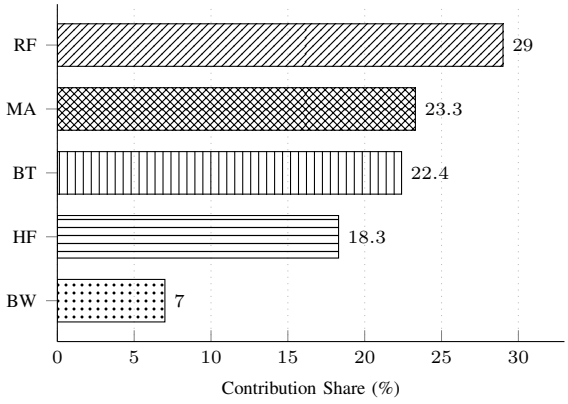

Fig. 1. Shapley Shares Under the Selected Internal-Label Utility

Table III lists the highest internal-label-retention combination at each size. RF alone retains 37.4%; adding MA reaches 59.6%; adding HF reaches 78.0%; adding BT reaches 95.2%; all five retain 100%.

TABLE III
HIGHEST INTERNAL-LABEL RETENTION AT EACH SIZE

k	**Best Combination**	**Internal-Label Retention**
1	RF	37.4%
2	RF+MA	59.6%
3	HF+RF+MA	78.0%
4	BT+HF+RF+MA	95.2%
5	BT+BW+HF+RF+MA	100.0%

At the address-project-record level, 78.0% of the 964,034 internally flagged records trigger exactly one indicator, 19.0% trigger two, and 3.0% trigger three or more. This distribution shows sparse simultaneous rule activation at the record level; it does not identify unique human actors or establish how many real-world techniques an actor uses. This also explains why HF exceeds BT in the highest-retention three-indicator set (despite BT having a higher Shapley value): the records retained by BT are mostly already retained by RF and MA, while HF adds a different group. For operators, fund-flow indicators have relatively large marginal shares under the internal utility; deployment priority requires separate campaign-specific validation.

B. Internal-Label Retention Under Different Simultaneous-Trigger Requirements

For the same HF+RF+MA configuration, OR-1 retains 78.0% of the five-indicator internal label, whereas OR-2 retains 10.3% and OR-3 retains 0.3%. The decline shows that simultaneous activation of multiple selected indicators is uncommon at the address-project-record level. It does not determine which rule provides the best false-positive–false-negative trade-off against independently verified ground truth.

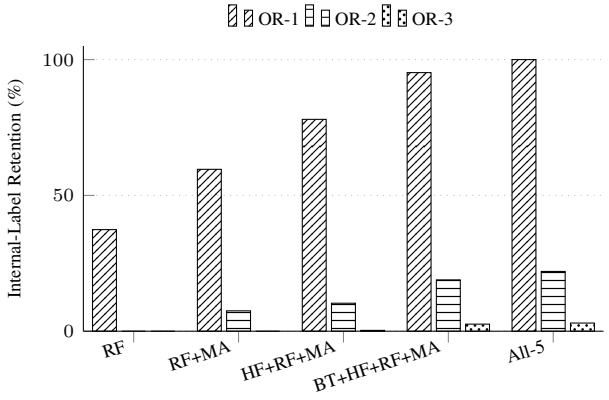

Fig. 2. Internal-Label Retention Under Different Simultaneous-Trigger Requirements

Within this selected address set, the largest-magnitude pairwise correlation is between HF and RF ($r = -0.48$). This conditional address-level statistic differs from the address-project-record unit used for retention and is not independent evidence of behavioral complementarity, distinct hunter types, or actor strategies.

From a practical standpoint, the three-indicator OR-1 combination (HF+RF+MA) is the highest-retention three-rule subset under the current internal label and may be considered as a candidate screening configuration for human review; the present analysis does not establish it as an optimal deployment policy.

From a decision-support perspective, a single-trigger rule is structurally simpler because each flag can be associated with one explicit rule condition. Whether this improves operator comprehension, auditability, or decision performance requires human-subject evaluation.

C. Agreement with External Flagged-Address Lists

After address normalization (standardizing format and removing duplicates), Gitcoin has 27,984 addresses, Hop has 14,195, and LayerZero has 21,395. They barely overlap: 99.4% of addresses appear in only one list.

TABLE IV
AGREEMENT WITH EXTERNAL FLAGGED-ADDRESS LISTS (OR-1)

Source	**List–HasciDB Intersection**	**RF**	**HF+RF+MA**	**All Five**
Gitcoin	8,856	19.4%	26.4%	37.6%
Hop	2,672	42.0%	49.2%	58.6%
LayerZero	495	25.5%	27.5%	33.5%

The three rule sets are fixed from the internal analysis and are not re-optimized for any external list. Among addresses appearing in both an external list and HasciDB, the all-five

rule agrees with 58.6% of the Hop intersection, 37.6% of the Gitcoin intersection, and 33.5% of the LayerZero intersection. These percentages use different external intersection denominators and have no verified negative class.

An important clarification: the internal 78–95% values and external 33.5–58.6% values have different denominators and meanings. Internal retention asks what fraction of five-rule OR-positive address-project records a subset preserves. External agreement asks what fraction of externally listed addresses that occur in HasciDB have triggered the selected rules in their available HasciDB histories. The two quantities should not be compared as detector-performance scores.

For operators, the Address-list agreement provides descriptive evidence, not proof of policy portability; campaign-specific validation remains necessary.

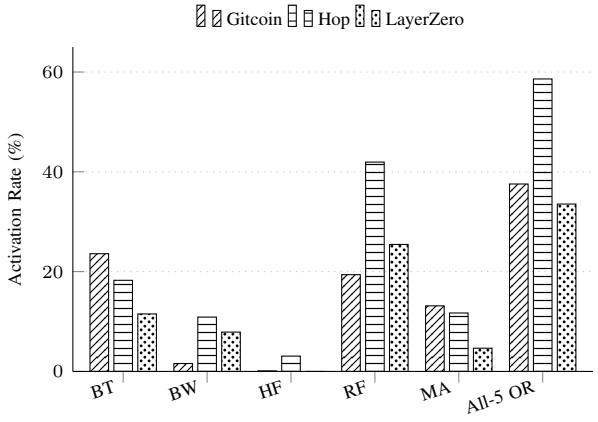

Fig. 3. HasciDB-History Indicator Activations Among Externally Listed Addresses

The externally listed address groups exhibit different indicator-activation profiles in their available HasciDB histories (Fig. 3). These differences may reflect population composition, list-construction criteria, HasciDB observation coverage, or project-related behavior. Because indicators were not recomputed within each external campaign and the analysis is cross-sectional, the result does not establish that the same actors adapted their strategies across projects.

D. Dominant Rule-Trigger Indicators Across Sectors

Table V shows which indicator has the highest exclusive trigger rate in each project's internally flagged address-project records. Trading projects are dominated by BT and RF; NFT projects show diverse patterns (MA, HF, RF); Staking projects favor MA; Infrastructure/Governance projects favor RF and MA. The overall ranking (RF $>$ MA $>$ BT $>$ HF $>$ BW) is descriptive rather than universal; these project-level trigger distributions do not identify actor techniques or causal sector effects.

Among Hop-listed addresses, RF and BW are relatively frequent in their available HasciDB histories. Because these activations were not necessarily produced within the Hop campaign, they should not be interpreted as direct evidence about behavior in Hop itself. The same limitation applies to the Gitcoin and LayerZero listed-address profiles.

TABLE V
MOST PROMINENT INDICATOR PER PROJECT (RATE = FRACTION OF THAT PROJECT'S INTERNALLY FLAGGED ADDRESS-PROJECT RECORDS THAT TRIGGER *only* THIS INDICATOR AND NO OTHERS)

Sector	Project	Dominant	Rate
Trading	1inch	BT	44.9%
	ParaSwap	BT	62.1%
	Uniswap	RF	48.8%
	dYdX	RF	64.8%
NFT	LooksRare	MA	45.9%
	PENGU	MA	37.3%
	X2Y2	HF	36.1%
	Apecoin	RF	47.0%
	Blur S2	RF	40.5%
	Blur S1	RF	27.1%
Staking	EigenLayer	MA	49.1%
	Badger	MA	43.4%
	Ether.fi	RF	55.9%
Infra/Gov	ENS	RF	54.4%
	Gitcoin	RF	44.9%
	Ampleforth	MA	47.5%

In policy terms, operators should read indicator rankings as descriptive project-level trigger summaries, not verified ecosystem behavior profiles.

E. Exploratory Analysis of Unmatched Hop-List Addresses

Of the Hop intersection addresses, 41.4% are not flagged by the five-rule OR definition. We examined these 1,106 missed addresses: 22.0% have no nonzero values among the five stored HasciDB indicators in their available records. They cannot be recovered by changing thresholds for these five indicators alone, although other data or features may still identify relevant activity. The remaining 78.0% have at least one nonzero value but remain below all current thresholds. Under the normalized value-to-threshold ratio used here, HF is nearest for 61.0%, BW for 25.0%, MA for 12.3%, and BT or RF for 1.7%.

This descriptive diagnostic identifies HF as a candidate for Hop-list-specific sensitivity analysis. It does not establish an optimal threshold reduction, the number of addresses recovered under a particular change, or the associated true false-positive cost.

F. Candidate Configurations and Evidence Boundaries

Table VI summarizes internal-retention results, candidate configurations, and deployment conditions that remain unvalidated. It is a decision-support summary rather than a recommendation of one universal policy.

Two practical implications follow. First, OR-1 should be treated as a screening and triage rule, not as an irreversible exclusion mechanism in every context. Because our internal labels are constructed from the same indicators, we cannot directly estimate false-positive rates on legitimate borderline users from HasciDB alone. This matters especially for MA in NFT-like markets, where normal buy-sell loops can resemble suspicious circular behavior. In deployment, operators

TABLE VI
CANDIDATE CONFIGURATIONS AND EVIDENCE AVAILABLE WITHIN THIS STUDY

Operator Goal	Candidate Configuration	Evidence Within This Study
Fewest-rule configuration	RF with OR-1	Highest single-rule internal-label retention (37.4%); actual monitoring cost is not measured.
Three-rule retention candidate	HF+RF+MA with OR-1	Highest three-rule internal-label retention (78.0%); not an established balanced or optimal default.
Higher internal-label retention	BT+HF+RF+MA with OR-1	Retains 95.2% of the parent internal label with four rules.
Cross-ecosystem reuse consideration	Validate with campaign-relevant data before selection	External listed-address groups differ, but the present analysis does not validate policy transfer.
Explore unmatched Hop-listed addresses	HF sensitivity-analysis candidate	HF is nearest for 61.0% of nonzero-but-below-threshold addresses in the Hop–HasciDB intersection.

should therefore pair OR-1 with project-specific allowlists, manual review for high-value accounts, or an appeal path when exclusion has material financial consequences. Second, threshold maintenance should be periodic rather than static. As a conservative operational heuristic, operators may review thresholds before each new campaign and after material changes to reward rules or relevant audit evidence. The present static analysis does not determine an optimal review frequency.

VI. DISCUSSION AND LIMITATIONS

This study provides an operator-readable rule-design artifact: indicator rankings, rule-size tradeoffs, and candidate configurations. Because exclusions have financial consequences, OR-1 rules should support screening and review rather than automatic exclusion; campaign-relevant validation, allowlists, or appeals may be needed. Whether these artifacts improve comprehension or decisions is untested.

The principal evidence boundary is that HasciDB's positive label is constructed from the same five indicators. Thus, internal results measure retention, overlap, and utility-based attribution, not independently verified detection accuracy. External results likewise measure agreement in available HasciDB histories, not recovery of the same external behavior; the lists represent different populations and their intersections, especially LayerZero's, are limited.

The indicators also have blind spots: RF omits indirect consolidation, MA can capture legitimate NFT-market loops, and HF's 180-day cap changes its meaning for long-window projects. The static analysis neither estimates false-positive rates nor establishes threshold sensitivity, strategic adaptation, or an optimal refresh schedule. These questions require independently labeled, longitudinal, and human-subject studies.

VII. CONCLUSION

Using five interpretable indicators and 2.5 million address-project records, we find that RF and MA have the largest Shapley shares under the selected internal-label utility, single-trigger rules retain more of the parent OR label than multi-trigger rules, and external lists have different HasciDB-history profiles. These results offer an auditable basis for documenting candidate screening rules, not a validated universal detector or policy.

REFERENCES

[1] C. Li, H. Chen, and W. Cai, "From slang to standards: Consensus-driven airdrop hunter definition as a baseline for cryptocurrency ecosystem security and governance," in *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, ser. CHI '26. New York, NY, USA: ACM, 2026.

[2] Z. Liu and H. Zhu, "Fighting Sybils in airdrops," *arXiv preprint arXiv:2209.04603*, 2022.

[3] C. Zhou, H. Chen, H. Wu, J. Zhang, and W. Cai, "Artemis: Detecting airdrop hunters in NFT markets with a graph learning system," in *Proceedings of the ACM Web Conference 2024 (WWW '24)*. ACM, 2024, pp. 1824–1834.

[4] Y. Qin, T. Ma, H. Chen, and H. Duan, "ARTEMIX: A community-boosting-based framework for airdrop hunter detection in the Web3 community," *Blockchain*, vol. 2, no. 2, p. 0010, 2024.

[5] S. Meiklejohn *et al.*, "A fistful of Bitcoins: Characterizing payments among men with no names," in *Proceedings of the ACM Internet Measurement Conference (IMC)*. ACM, 2013, pp. 127–140.

[6] D. Ron and A. Shamir, "Quantitative analysis of the full Bitcoin transaction graph," in *Financial Cryptography and Data Security (FC '13)*. Springer, 2013, pp. 6–24.

[7] H. Kalodner, M. Möser, K. Lee, S. Goldfeder, M. Plattner, A. Chator, and A. Narayanan, "BlockSci: Design and applications of a blockchain analysis platform," in *Proceedings of USENIX Security '20*. USENIX, 2020, pp. 2721–2738.

[8] T. Chen, Z. Li, Y. Zhu, J. Chen, X. Luo, J. C.-S. Lui, X. Lin, and X. Zhang, "Understanding Ethereum via graph analysis," *ACM Transactions on Internet Technology*, vol. 20, no. 2, pp. 1–32, 2020.

[9] Nansen, "On-chain analytics platform," https://www.nansen.ai, 2024, accessed: March 2026.

[10] TrustaLabs, "On-chain trust and safety," https://trustalabs.ai, 2024, accessed: March 2026.

[11] Human Technology, "Human passport: Proof of personhood and Sybil resistance for Web3," https://human.tech, 2024, acquired from Gitcoin Passport, protecting $430M+ across 120+ projects.

[12] Worldcoin Foundation, "World ID: Privacy-preserving proof of personhood," https://whitepaper.world.org, 2025, 16M+ verified users, 1,500+ Orbs in 23 countries.

[13] J. R. Douceur, "The Sybil attack," in *International Workshop on Peer-to-Peer Systems (IPTPS)*. Springer, 2002, pp. 251–260.

[14] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman, "SybilGuard: Defending against Sybil attacks via social networks," in *Proceedings of ACM SIGCOMM '06*. ACM, 2006, pp. 267–278.

[15] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao, "SybilLimit: A near-optimal social network defense against Sybil attacks," in *Proceedings of IEEE S&P '08*. IEEE, 2008, pp. 3–17.

[16] Gitcoin, "Fraud detection & defense (FDD): Sybil address dataset," Gitcoin Grants Program, 2024, sAD model predictions from GR15. [Online]. Available: https://gov.gitcoin.co/t/fdd-review

[17] Hop Protocol, "Hop airdrop distribution," GitHub repository, 2022. [Online]. Available: https://github.com/hop-protocol/hop-airdrop

[18] LayerZero Labs, "Addressing Sybil activity," Medium, 2024. [Online]. Available: https://medium.com/layerzero-official/addressing-sybil-activity