Poster: From TOR to ORB28: Towards Formal Anonymity Guarantees

Sunkanmi Oluwadare
School of Information Technology
University of Cincinnati
Email: oluwadse@mail.uc.edu

Jacques Bou Abdo
School of Information Technology
University of Cincinnati
Email: bouabdjs@ucmail.uc.edu

1. Introduction

The Tor network has been studied extensively, yet a critical analytical foundation remains missing. Murdoch and Zielinski [1] sampled traffic at Internet Exchange Points and showed that a small number of IXPs carry a disproportionate fraction of Tor circuits, revealing structural centralization at the infrastructure layer. Edman and Syverson [2] pushed further, demonstrating that Autonomous System-level overlap allows a single AS to observe both ends of a significant fraction of circuits, and proposed AS-aware relay selection heuristics to mitigate this.

Critically, the topology these studies examine is physical and geographic: AS memberships, IXP positions, and physical locations. The topology this work examines is different: the emergent functional structure that arises when millions of independent clients simultaneously use Tor, a **systems science** concept [3] whose properties directly determine path diversity, structural centralization, and the anonymity guarantee each user actually receives.

Johnson et al. [4] used TorPS to show that traffic correlation attacks succeed at alarming rates when adversaries control relays or occupy strategic networks over time. McCoy et al. [5] operated Tor relays directly, documenting skewed bandwidth consumption and uneven exit relay utilization. Greubel et al.[6] quantified measurement quality and confirmed that Tor's bandwidth-weighted selection concentrates circuits on a small subset of high-capacity nodes.

What all of these studies share is a common limitation: without a formal model of how the selection algorithm produces the network's structure, the anonymity guarantee for any individual user remains unknown. Prior work can tell you that centralization exists; it cannot tell you what that centralization means for the user routed through a low-bandwidth circuit versus the user routed through an elite hub. It cannot bound the best-case or worst-case anonymity that the network structurally permits. And it cannot prove whether a well-resourced adversary can exploit that structure to compromise a disproportionate share of users. These are not observational questions. They are structural ones, and they require a formal model to answer.

This gap cannot be closed observationally. Measurement studies document that centralization exists but cannot predict how it shifts under protocol changes or adversarial conditions, and cannot quantify the anonymity afforded to each user. Security claims about Tor therefore rest on empirical snapshots rather than validated theory.

Recent operational intelligence makes this gap urgent. China-nexus MURKY PANDA deployed the newly created ORB network ORB28 specifically for anonymity, while Russia-nexus COZY BEAR abused authentication flows for account compromise [7], both exploiting the structural opacity that emerges when anonymity networks are never formally characterized. These networks are engineered to defeat traffic observation, AS-level path inference, and relay compromise detection, the exact methods Tor's security literature relies upon [8]. Unlike indicators of compromise, which adversaries actively rotate, structural properties are persistent. A formal model of how anonymity networks form and concentrate traffic is therefore not an academic exercise; it is the prerequisite for any defense that can reason about adversarial positioning before the adversary acts. We address this with a two-phase formal methods approach: deriving and validating the mathematical structure of Tor's emergent topology through analysis of its relay selection modules, then using that model to measure structural anonymity guarantees.

2. Methodology

The approach used in this study proceeds in two connected phases.

2.1. Formal Topology Model

1) **Source Code Analysis:** We analyzed Tor's C codebase, focusing on `routerlist.c`, `node_select.c`, and `circuitbuild.c`. We confirmed the implementation facts and also that the selection mechanism follows the weighted random sampling without replacement framework of Efraimidis and Spirakis [9].
2) **Mathematical Modeling:** Building on the source code, we formally characterized Tor's emergent network topology as a *tripartite weighted random multigraph*. The network consists of two bipartite weighted random multigraphs: MG_1 connecting Guard and Middle relays, and MG_2 connecting

Middle and Exit relays, with link probabilities derived from the bandwidth-weighted selection process. For a relay n_i operating in role r, its expected degree, representing the number of concurrent circuits utilizing that node is:

$$\mathrm{Deg}(n_i) = N \times \frac{w_i}{\sum_{j \in R} w_j} \tag{1}$$

where N is the total number of active circuits and w_i is the relay's consensus bandwidth weight. The full model is developed formally in [10].

3) **Empirical Validation:** Validated against 88,121 live circuits over four weeks. Guard relays are excluded due to stateful client-side pinning. Source code analysis confirms Tor uses weighted random sampling without replacement, then our model validated by observed live circuit behavior.

2.2. Structural Anonymity Quantification

With the validated topology model [10] as our foundation, we operationalize k-anonymity as a function of the emergent network structure. The structural anonymity set size $\bar{K}$ for a typical circuit is derived as:

$$\bar{K} = N^3 \left(\sum_{n_i \in G} \omega_i^2 \right) \times \left(\sum_{n_j \in M} \omega_j^2 \right) \times \left(\sum_{n_k \in E} \omega_k^2 \right) \tag{2}$$

where ω_i is the normalized consensus bandwidth weight of relay n_i. Structural bounds K_{max} and K_{min} are derived analytically, adversarial compromise modeled via trust coefficient C_n, and validated through a Monte Carlo simulation of 10^5 circuits on live consensus data (January 23, 2026) as shown in figure 1.

3. Empirical Validation

The KS test results for Middle and Exit roles across the three datasets are summarized in Table 3.

Table 1. KS TEST RESULTS CONFIRMING MODEL FIDELITY (ALL P ¿ 0.05)

Dataset	Role	p-value	Interpretation ($\alpha = 0.05$)
tor_data (0)	Middle	0.4335	Not significant ($p > 0.05$)
tor_data (0)	Exit	0.1901	Not significant ($p > 0.05$)
tor_data (1)	Middle	0.1383	Not significant ($p > 0.05$)
tor_data (1)	Exit	0.4335	Not significant ($p > 0.05$)
tor_data (2)	Middle	0.4399	Not significant ($p > 0.05$)
tor_data (2)	Exit	0.1901	Not significant ($p > 0.05$)

4. Discussion & Conclusion

The central contribution of our work is a formal methods approach to anonymity guarantees in Tor. Prior studies have observed that centralization exists [1], measured its consequences at the infrastructure layer [2], modeled adversarial positioning [4], documented application-layer usage patterns [5], and approximated path selection behavior through simulation [4]. What none could do is prove, from first principles, what the network's structure means for any individual user, not as an average, but as a formal, derivable property of the topology. This work does exactly that.

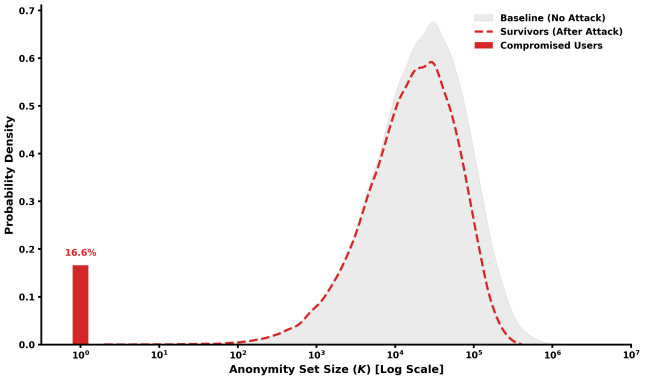


Figure 1. **Impact of Targeted Attack**. Structural anonymity distribution under targeted adversarial attack. Under honest conditions, the typical user sits at $\bar{K} \approx 44{,}358$ (mean) and $\tilde{K} \approx 22{,}468$ (median), safely above the unity barrier. When the top 1% of relays by bandwidth are compromised, 16.58% of circuits collapse directly to $K_{min} = 1$ (red bar) while survivors retain $K \approx 10^4$, with $K_{max} \approx 4.2 \times 10^6$ representing the theoretical ceiling.

References

[1] S. J. Murdoch and P. Zieliński, "Sampled traffic analysis by internet-exchange-level adversaries," in *Privacy Enhancing Technologies (PETS '07), LNCS 4776*. Springer, 2007, pp. 167–183.

[2] M. Edman and P. Syverson, "As-awareness in tor path selection," in *ACM Conference on Computer and Communications Security (CCS '09)*, 2009, pp. 380–389.

[3] J. H. Holland, *Emergence: From chaos to order*. OUP Oxford, 2000.

[4] A. Johnson, C. Wacek, R. Jansen, M. Sherr, and P. Syverson, "Users get routed: Traffic correlation on tor by realistic adversaries," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, 2013, pp. 337–348.

[5] D. McCoy, K. Bauer, D. Grunwald, T. Kohno, and D. Sicker, "Shining light in dark places: Understanding the tor network," in *Privacy Enhancing Technologies (PETS '08), LNCS 5134*. Springer, 2008, pp. 63–76.

[6] A. Greubel, S. Pohl, and S. Kounev, "Quantifying measurement quality and load distribution in tor," in *Annual Computer Security Applications Conference (ACSAC '20)*, 2020, pp. 1–14.

[7] CrowdStrike, "2026 Global Threat Report," CrowdStrike, Tech. Rep., feb 2026. [Online]. Available: https://crowdstrike.com/global-threat-report/

[8] M. Raggi, "IOC extinction? China-nexus cyber espionage actors use ORB networks to raise cost on defenders," May 2024, accessed: March 2026.

[9] P. S. Efraimidis and P. G. Spirakis, "Weighted random sampling with a reservoir," *Information processing letters*, vol. 97, no. 5, pp. 181–185, 2006.

[10] S. Oluwadare, J. B. Abdo, and M. Nassar, "Mapping the tor network: Modeling and empirical validation," *Authorea Preprints*, 2025.

From TOR to ORB28: Towards Formal Anonymity Guarantees.

School of Information Technology, University of Cincinnati.

Sunkanmi Oluwadare • Jacques Bou Abdo.

1. Introduction & Motivation.

The Tor network routes anonymous traffic through a three-hop circuit of Guard, Middle, and Exit relays selected by a bandwidth-weighted algorithm. Prior work has measured centralization, modeled AS-level adversaries, documented application-layer usage patterns, and simulated path selection but without a formal model, the anonymity guarantee for any individual user remains unknown.

The Gap: Prior studies examine physical/geographic topology: AS memberships, IXP positions. None derives the emergent functional structure that arises from bandwidth-weighted selection, the structure that directly determines each user's anonymity.

2. Methodology - Two-Phase Formal Approach.

Phase 1: Formal Topology Model

Source Code Analysis: Tor's C codebase - routerlist.c, node_select.c, circuitbuild.c - confirms relay selection follows Efraimidis-Spirakis weighted random sampling without replacement.

Mathematical Model: Tor's emergent network formally characterized as a tripartite weighted random multigraph (Mg_1 Guard↔Middle, MG_2 Middle↔Exit). Link probabilities derived from bandwidth-weighted selection.

$$\mathrm{Deg}(n_i) = N \times w_i / \Sigma_{j \in R} w_j$$

Empirical Validation: Model validated against 88,121 live circuits over 4 weeks across 3 datasets. KS tests confirm fidelity (all $p > 0.05$). Guard relays excluded due to stateful pinning.

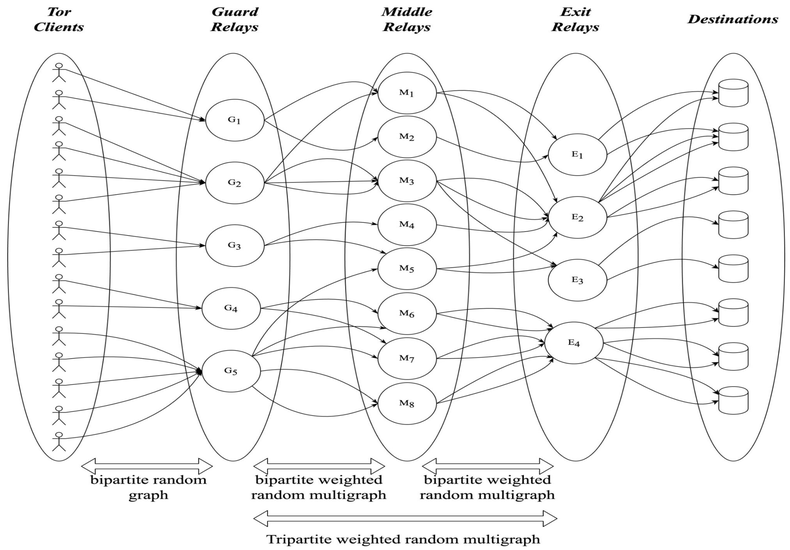
Phase 2: Structural Anonymity Quantification

With the validated topology as foundation, k-anonymity is operationalized as a structural function — anonymity derived, not observed:

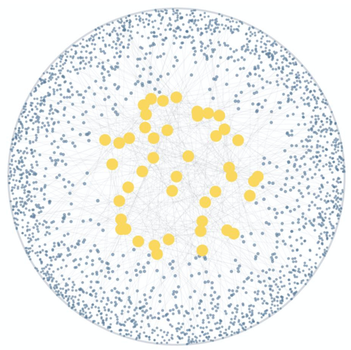
$$\bar{K} = N^3 \times (\Sigma\omega_i^2)G \times (\Sigma\omega_j^2)M \times (\Sigma\omega_k^2)E$$

3. Tor's Emergent Network: Tripartite Weighted Random Multigraph.

Emergent topology: the functional structure arising when millions of clients simultaneously execute bandwidth-weighted selection — a systems science concept.



Millions of independent clients simultaneously executing bandwidth-weighted selection give rise to this predictable, formally characterizable global structure: not planned, but emergent.



Top 1% of relays by bandwidth (yellow) dominate circuit routing, the emergent topology concentrates trust in a small structural core before any adversary acts.

4. Empirical Validation.

Dataset	Role	p-value	Result
tor_data(0)	Middle	0.4335	✓ p > 0.05
tor_data(0)	Exit	0.1901	✓ p > 0.05
tor_data(1)	Middle	0.1383	✓ p > 0.05
tor_data(1)	Exit	0.4335	✓ p > 0.05
tor_data(2)	Middle	0.4399	✓ p > 0.05
tor_data(2)	Exit	0.1901	✓ p > 0.05

Middle + Exit: stateless, bandwidth-proportional ✓
Guard: stateful, long-term pinned — excluded from KS test

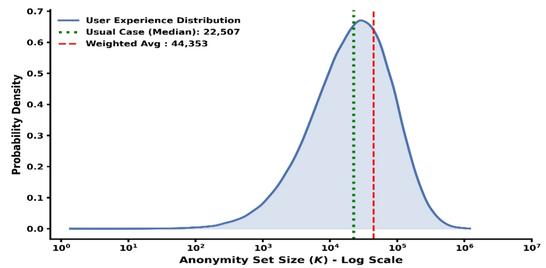
Key Results.

88,121 Live circuits validated

16:1 Adversarial leverage ratio

16.58% Circuits hit by top 1% attack

0.00% Circuits hit by bottom 1% attack

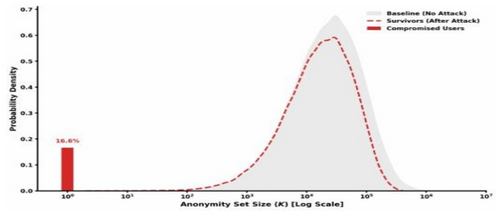


Glass Cannon Topology

$\bar{K} \approx 44{,}353$ (mean) $K \approx 22{,}507$ (median)
$K_{max} \approx 4.2 \times 10^6$ $K_{min} = 1$
Mean is nearly double the median - average metrics systematically overstate typical user protection. Structural range spans 6 orders of magnitude.

Users fall directly from $K = 10^4$ to $K = 1$

5. Adversarial Stress Test.



Top 1% targeted → 16.58% of circuits collapse to K=1 (red bar). Bottom 1% → 0.00%. Anonymity loss is binary.

6. Discussion & Conclusion.

Prior studies observed centralization exists , measured it, modeled adversarial consequences, simulated path selection. What none could do is prove, from first principles, what the network's structure means for any individual user.

This work establishes that anonymity in Tor is not a uniform guarantee , it is a structural variable, unequally distributed, formally quantifiable, and directly determined by the bandwidth-weighted algorithm that makes Tor fast.

Our Contributions:

- First code-derived, validated model of Tor's emergent topology.
- Anonymity quantification using K-anonymity as a function of the model.
- Glass Cannon: $\bar{K} \approx 44K$, $K_{min}=1$, $K_{max} \approx 4.2 \times 10^6$, 16:1 leverage.
- Anonymity loss proven binary — not gradual.

oluwadse@mail.uc.edu