New Jersey Voter ID Numbers Reconfigured with Shift Cipher

A Paquette

Independent Researcher

paqart@gmail.com

Abstract: *This study examines New Jersey voter identification numbers for hidden algorithmic attributes, following similar findings in New York, where algorithms could track certain illegal records. Analysis reveals complex obfuscation patterns in New Jersey, including shift ciphers, in these public-facing numbers. The purpose of such obfuscation in accessible data is unclear, raising transparency concerns. Approximately 102,854 potentially erroneous or falsified records were discovered, indicating serious integrity issues. While no direct link was found between these problematic records and the algorithmic modifications, their coexistence is concerning. The ability to algorithmically track illegal records, as seen in New York, suggests a potential for misuse. These findings have significant implications for election integrity and data management practices.*

Keywords: *New Jersey, Voter Rolls, Algorithm, Cipher, Election*

Introduction

The integrity of voter rolls is fundamental to the democratic process, serving as the bedrock of fair and transparent elections. Recent investigations into voter identification systems have revealed complex algorithmic patterns that raise important questions about data management, privacy, and the potential for manipulation in electoral systems.

Voter rolls are "critical infrastructure" as declared in 2017 by Secretary of Homeland Security Jeh Johnson (Hodgson *et al.* 2020, p. 2). The presence of obfuscation techniques within a database known to contain problematic (false) data suggests the possibility that the database has been compromised in unknown ways. A compromised voter roll database is compromised infrastructure. This can be dangerous, as explained by Kanuck: "inaccurate data undermines the value of information, impairs the reliability of information sources, and/or erodes public trust and confidence"(Kanuck 2022, p. 187).

Voter roll manipulation is a critical component of election fraud schemes, with effects that can subvert legitimate government control, similar to conventional warfare (Downes 2018). As stated by Baptist & Gluck 2021, (pp. 40), "one can sway democratic states merely by influencing the citizenry—or its officials—to vote (or not vote, as the case may be) in a specific way favourable to a foreign entity". Fake registrations can generate or reconcile fraudulent ballots, necessitating sophisticated obfuscation in public voter rolls. Large-scale fraud requires false registrations to be both hidden and accessible to conspirators, a challenge compounded by the public nature of voter rolls. The algorithms found in New York and New Jersey could potentially be used for such purposes, although it is not possible to determine if they were designed or used nefariously from available data. Regardless, their presence in the voter rolls warrants attention and scrutiny.

This study examines voter identification systems and their vulnerabilities from a non-partisan, data-driven perspective. It focuses on technical aspects, disregarding political affiliations. Voter rolls, while containing party information, are analysed objectively as they don't link registrations to specific ballots or dictate voting choices.

A recent study of New York State Board of Election voter rolls revealed complex algorithmic patterns in voter ID numbers (Paquette 2023). These patterns mapped county to state voter IDs, effectively tagging records with hidden attributes that could predict certain suspicious records. This discovery prompted an examination of New Jersey's voter rolls to determine if similar algorithms could disclose hidden attributes. Multiple algorithms assign voter ID numbers across New Jersey's 21 counties, similar to New York's "Spiral" and "Tartan" algorithms. However, New Jersey's algorithms alter ID numbers, unlike New York's mapping approach.

ID number transformation can protect sensitive information and can help maintain research objectivity. However, it's not foolproof. For instance, when New York City released anonymized taxi data, malicious actors reidentified drivers, fares, and destinations, enabling potential extortion based on embarrassing inferences (Goroff *et al.* 2018). Disclosure risk only applies to private data, not publicly available information where disclosure is certain. Yet, New York and New Jersey have algorithmically modified voter ID numbers or their relationships, seemingly to add noise—a technique typically used to obfuscate sensitive data (Templ & Sariyar, 2022). This implies the presence of risky data in the voter rolls, although its nature is unclear. While county Boards of Elections may use driver's license or partial SSN numbers for identification, these are not included in public voter rolls.

The Voter Registration Act of 1993 mandates public access to voter roll records for transparency (National Voter Registration Act 1993). Unlike the New York taxi data, which required additional information for reidentification, voter rolls explicitly link ID numbers to individuals' names, birthdates, and addresses. Thus, manipulating ID numbers in voter rolls cannot conceal the direct mapping between numbers and individuals. For example, if ID number 12,345,678 is connected to John Doe at 123 Maple Street, changing the ID to AU8761 has no effect on this mapping; the modified number remains linked to the same information, rendering such obfuscation ineffective for privacy protection or for security in this public dataset.

In the Supreme Court decision Ex Parte Coy (1888), a commonsense interpretation of law governed the outcome of a case concerning election fraud. Specifically, a criminal conspiracy was hatched to induce election officials to not perform their duties so that a local candidate in Indiana would win his election. According to the majority opinion, "Crimes against the ballot have become so numerous and so serious that the attention of all legislative bodies has been turned with anxious solicitude to the means of preventing them" (Ex Parte Coy 1888). They go on to declare that lack of intent to cause specific harm (or benefit) to a named individual is no defence for negligence in matters involving the integrity of elections.

The court's position was that, although the appellants conspired to commit fraud in a small election, their actions compromised all other elections in the region. Therefore, the potential and actual harm of their actions to unintended victims, though absent criminal intent, had to be punished. Concealment of voter roll information by obfuscation methods, whatever their purpose, violates the stated purpose of public disclosure laws. The Coy decision is consistent with the position that obfuscation of voter roll data causes harm to the integrity of the electoral process, and thus to every affected voter, regardless of intent.

No correlation between irregular registrations and voter ID algorithms was found in New Jersey's rolls, unlike in New York. However, this may be due to limited data availability rather than absence of such correlations. While algorithms similar to New York's exist in New Jersey, their purpose remains unclear. Privacy or security justifications are implausible, given the public nature of all fields. The complexity of these algorithms suggests deliberate design, but their function in public data management is not apparent.

The algorithms in New York and New Jersey create "shadow data"—information not defined by any existing database fields. This shadow data, as Balmer notes, signifies something about its source, even if that source is unknown (Balmer 2017). In the context of voter rolls, this creates a hidden attribute within ostensibly public data. This is at odds with the National Voting Rights Act's requirement for full public disclosure to ensure "currency and accuracy" of voter lists. The presence of disguised, purposefully opaque data within these public records raises questions about compliance with transparency mandates.

If the purpose of the voter ID algorithms was simply to assign ID numbers, the algorithms are over-engineered. Moreover, they make the system unnecessarily complex, less efficient, create novel database management requirements, and have no known connection to visible data fields in their respective databases. Each of these characteristics is inconsistent with database management norms. This paper will describe distinctive characteristics of New Jersey's voter roll algorithms and will discuss how they may have been used to conceal attributes of suspicious records.

Discovery Process and Analysis

This study aimed to identify suspicious records (potentially illegal) and to uncover any unusual voter ID-related algorithms in New Jersey's public voter rolls. The approach was exploratory and iterative, driven by pattern recognition and data anomalies rather than a predetermined protocol.

Unlike in the New York study, in this study, New Jersey's public voter rolls were found to lack data such as purged records, purge dates, last vote dates, and voter history. This limitation narrowed the focus to three main criteria for identifying suspicious records: "clones", impossible registration dates, and age discrepancies. While not exhaustive, these criteria provided sufficient data points to initiate correlation analysis if an algorithm was discovered.

The investigation began with a thorough examination of voter ID numbers, of which two types were present in the New Jersey database. The process involved:

1. Visual scanning of the data to identify potential patterns or anomalies;
2. Gap analysis on various sort methods to reveal potential relationships between ID numbers;
3. Grouping analysis to identify numbers that appeared to be related; and
4. Transformation analysis to uncover potential encoding or encryption methods.

This iterative process led to the discovery of multiple well-defined algorithms used in ID number assignment. The algorithms' complexity and consistency across the dataset strongly suggested deliberate implementation rather than random artifacts.

Suspicious records were categorized as follows:

- "Clones": Duplicate records with unique ID numbers
- Impossible registration dates: Earlier than the voter's 15th birthday or on holidays when registration tasks are unlikely to be performed
- Age discrepancies: Voters older than the oldest known living American in 2020 (115 years old)

While this approach may not have captured all suspicious records, it provided a robust foundation for identifying patterns and correlations between these records and the discovered algorithms.

Findings

Suspicious records

Analysis of the New Jersey voter rolls revealed 102,854 records that met the criteria for suspicion, as defined in this study's methodology (**Table 1**). These records exhibit characteristics that deviate from expected norms and warrant further scrutiny.

Flagged	Count
clones	18,067
<15 YO onRD	21,819
>115 YO on RD	28,351
Christmas Day RD	586
July 4 RD	1,402
New Years Day RD	31,836
Thanksgiving RD	793
Total	102,854

Table 1: Flagged records

Detailed investigation of clone records uncovered patterns that cannot be explained by routine administrative processes, such as inter-county moves (**Table 2**). Notably:

- 3,333 clones had identical addresses to their originals, affecting at least 6,666 records.
- 948 of these also shared registration dates, impacting at least 1,896 records.
- At least 13,770 clone records were simultaneously active.

The presence of identical addresses and registration dates in cloned records eliminates the possibility of these being legitimate records of voters who have moved. Moreover, these findings directly contradict federal law, specifically the Help America Vote Act of 2002 (Help America Vote Act, 2002 §303) , which mandates one "unique identifier" per voter. The existence of cloned records is not only suspicious, but also illegal, as they could enable multiple votes per voter.

	Names	Addresses	RegDate	Addr and RegDate	<15 YO on RegDate	Active	Inactive
Unique	8,962	14,734	17,088	17,119			
Excess	9,105	3,333	979	948		6,885	
Total	18,067	18,067	18,067	18,067	55	15,611	2,297

Table 2: Cloned record breakdown shows high numbers of simultaneously active records and voters who could not have been cloned due to a move

ID numbers

New Jersey's voter rolls use two ID numbers: Legacy ID (LID) and Display ID (DID). All 6,459,433 records have a DID, but only 88.95% have an LID. The distribution of registration dates suggests DIDs were introduced in 2019 and retroactively applied: only 0.12% of non-LID records predate 2019, compared to 11.99% of LID records postdating 2018. LID records span 1850-2022 (excluding 83 erroneous pre-1850 entries).

At least two algorithms are used in New Jersey's voter rolls. Understanding DID and LID relationships requires knowledge of their components, occupied state space, and governing algorithms.

DIDs have 11 characters: an "Alpha" (A-P) followed by a 10-digit "DID Num". The DID Num is analytically divided into "Left DID" (first five digits) and "Right DID" (last five digits). LIDs are nine-digit numbers, analytically divided into "Left LID" (first three digits) and "Right LID" (remaining six digits). These divisions are derived from analysis, not official nomenclature (**Figures 1 and 2**).

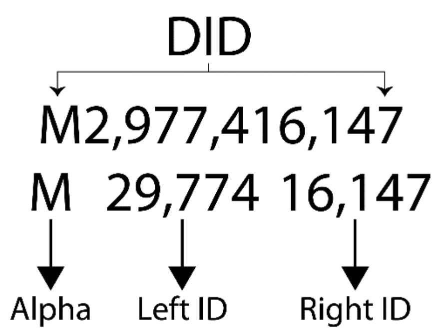

Figure 1: Format, New Jersey Display ID (DID) number

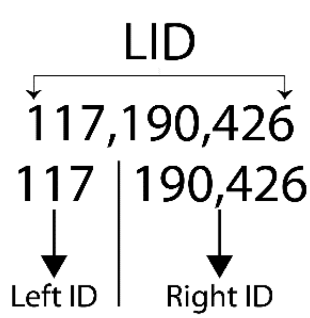

Figure 2: LID format, broken into Left LID and Right LID

New Jersey's ID number state space is defined by digit count: 10 for DIDs (max 9,999,999,999) and 9 for LIDs (max 999,999,999). Only 0.06% of possible DIDs are used due to the low record count. The state space is partitioned into three sub-partitions with varying assignment densities. Only 1.52% of records have DID numbers above 6,459,400,000, with the remaining 98.48% below, allowing one record per 1,015 DIDs for most of the database. Records without LID numbers (n=713,680) occupy the same DID state space as those with LIDs (**Figure 3**).

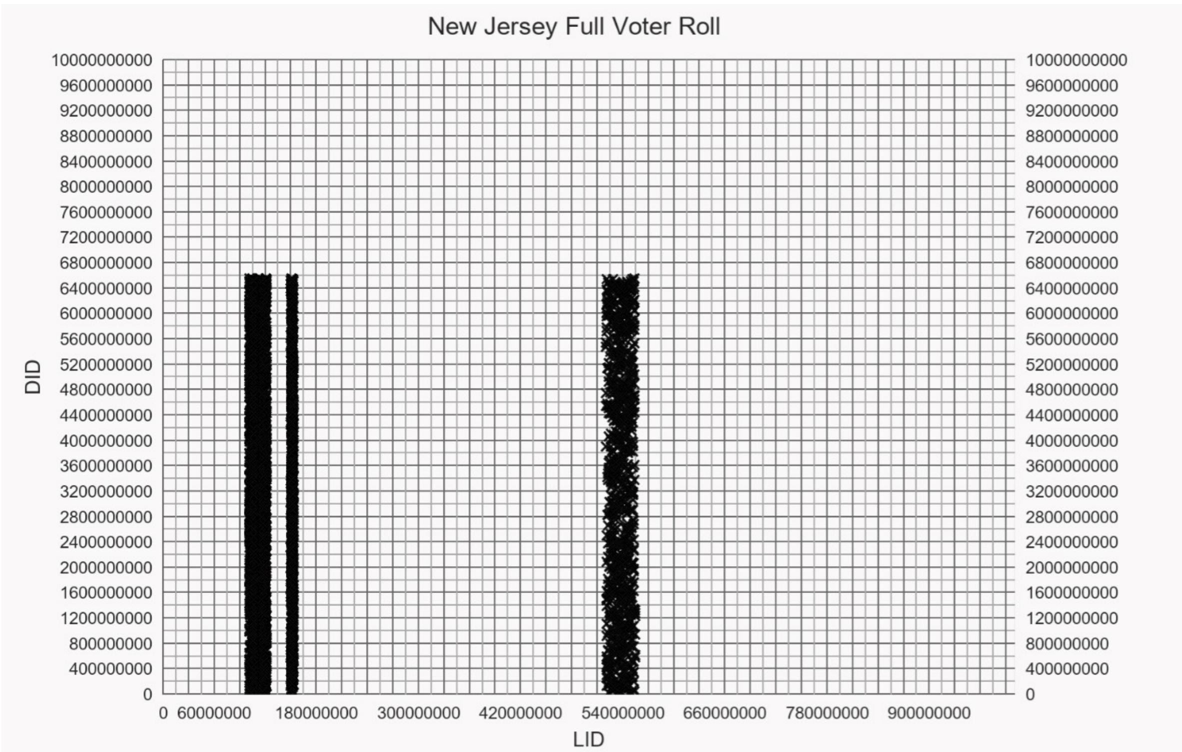

Figure 3: New Jersey state space and assigned numbers

When sorted by alpha character, New Jersey's DID numbers form linear patterns in a scatterplot (**Figure 4**). Each alpha character spans the full DID state space. Only 1,968 DID Num records have duplicate values with different alphas; all others are unique. This uniqueness is achieved through regular gaps in each alpha sequence to prevent collisions.

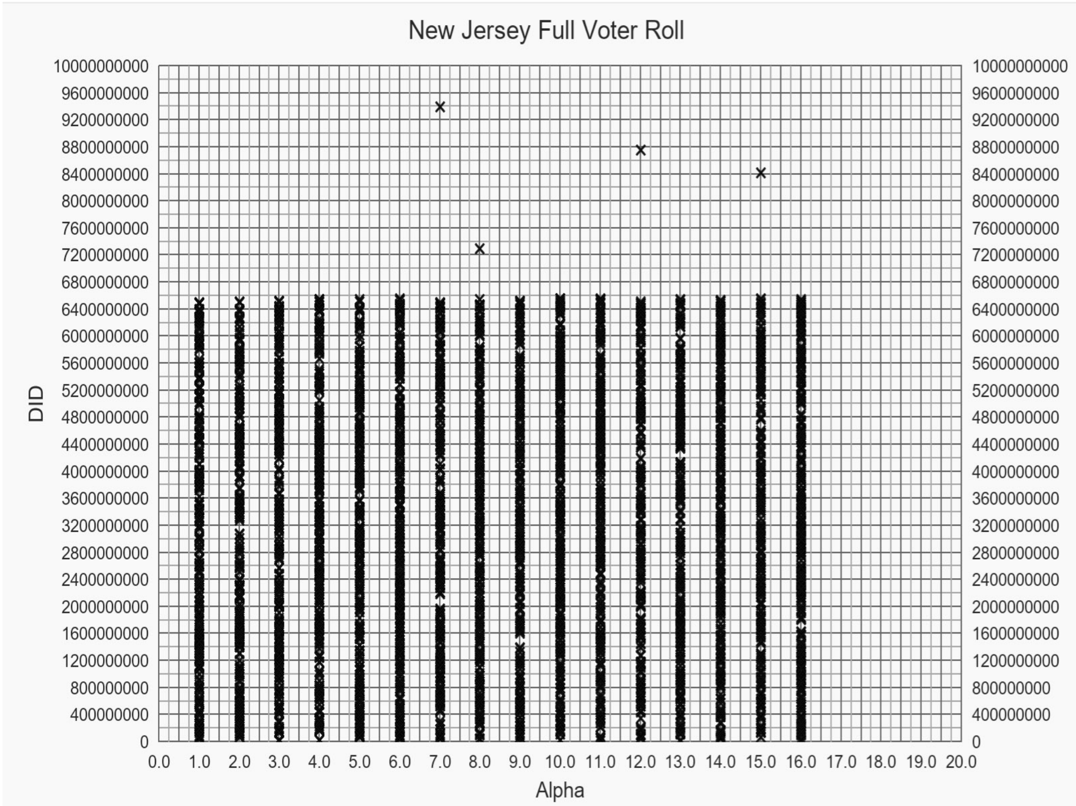

Figure 4: DID numbers, full New Jersey voter rolls, sorted by alpha

Algorithms

Two algorithms, "Arcade" and "Icicle", are identified in New Jersey's voter rolls, each occurring in three state space partitions (Arcade 1-3, Icicle 1-3; [**Table 3**]). Partitions have distinct characteristics affecting algorithm output. Icicle patterns only occur in records with LID numbers, while Arcade patterns span all assigned numbers. Pattern identification depends on: 1) LID presence (Icicle) or absence (Arcade), 2) Right DID (for Arcade) or LID (for Icicle) determines the pattern number.

Pattern	Min RID	Max RID	Min Leg_ID	Max Leg_ID	Leg_ID	Display ID	records	Numbering
Arcade 1	0	51,173	None	None	NO	YES	555,536	Consecutive
Icicle 1	**51,174**	**52,812**	**101,000,108**	**121,458,917**	**YES**	**YES**	**2,969,091**	Consecutive
Icicle2	**52,812**	**53,533**	**150,000,000**	**152,819,034**	**YES**	**YES**	**1,447,743**	Consecutive
Icicle 3	**53,533**	**54,073**	**520,000,000**	**554,166,601**	**YES**	**YES**	**1,328,919**	**Every 23**
Arcade 2	51,174	54,073	None	None	NO	YES	30,565	Consecutive
Arcade 3	54,074	96,584	None	None	NO	YES	127,579	Consecutive
Total							6,459,433	

Table 3: Partition ranges

New Jersey's Right DID state space spans 1-99,999 (**Table 4**). Icicle algorithm, used for 89.42% of records (5,776,317 out of 6,459,433), occupies only 2.9% (2,900) of available Right DID values.

Available Right DID numbers vary significantly by partition.

	Arcade 1	Icicle 1-3	Arcade 2	Arcade 3	Total
Start Right DID	1	51,174	54,074	65,536	
End Right DID	51,173	54,073	65,535	96,584	
Total numbers in range	51,173	2,900	11,462	31,049	96,584

Table 4: In-range and out of range partitions. Partition 2 uses the Icicle algorithm. Partitions 1, 3, and 4 use Arcade I, Arcade II, and Arcade III.

Records with Arcade ID numbers represent only 10.58% of all assigned numbers but utilize 97.1% of the available state space for Right DID numbers (**Figure 5**).

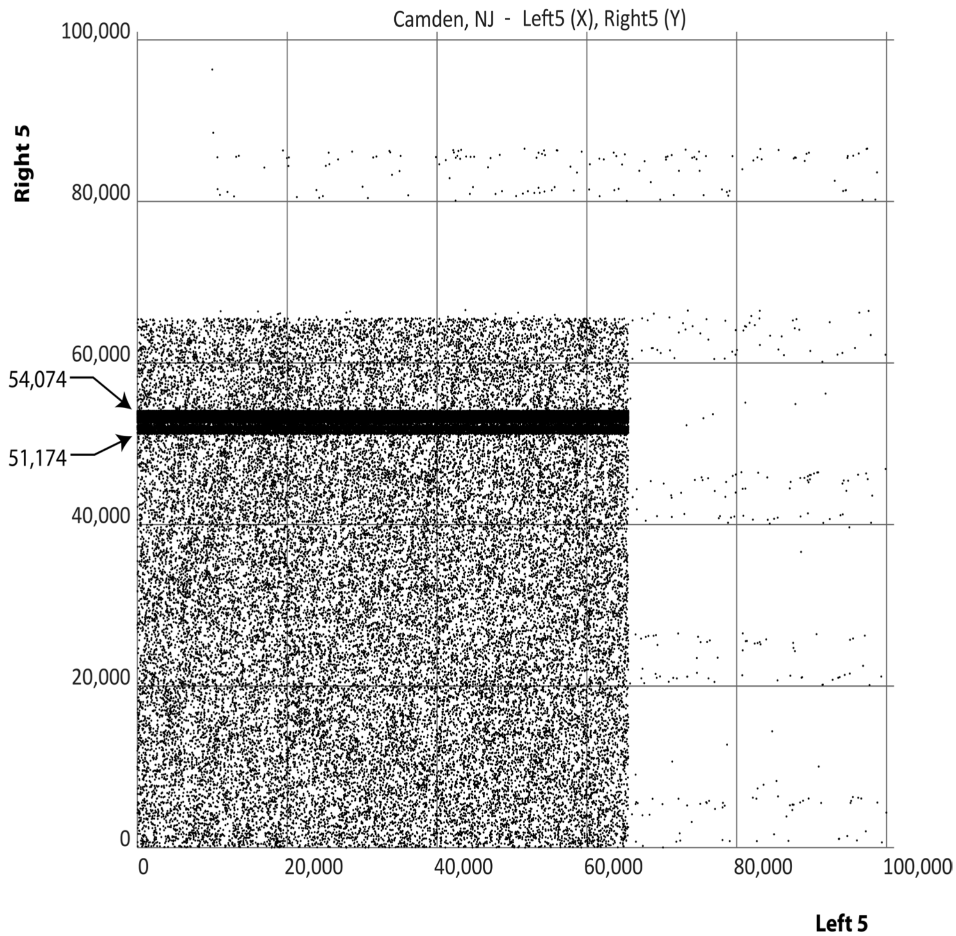

Figure 5: DID numbers split into Left ID (X-axis) and Right DID (Y-axis) reveal concentration of 10-digit numbers that end with the values 51,174-54,074 (Icicle pattern)

This distribution of numbers is unnatural. If numbers were assigned sequentially or randomly without bias, the resulting scatterplot would present an even distribution. Instead, there are well-defined stripes of different number densities within a larger area filled with sparse noise, as in **Figure 5**.

Arcade

Arcade patterns exhibit a pseudo-random distribution resembling the video game "Space Invaders" when Right DID (Y-axis) is plotted against assigned counties (X-axis [**Figure 6**]). Arcade 3, the least dense, uses the highest Right DID values. Arcade 2 intersperses with all three Icicle patterns. Arcade 1, the densest, has the lowest Right DID values statewide.

Right ID	Atlantic	Bergen	Burlington	Camden	Cape May	Cumberla	Essex	Glouceste	Hudson	Hunterdo	Mercer	Middlesex	Monmout	Morris	Ocean	Passaic	Salem	Somerset	Sussex	Union	Warren	All
66550	0	0	1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	5
66552	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
66553	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
68096	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
68113	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
68151	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
68163	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
68175	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
68197	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
68223	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
68326	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
68337	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
68347	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
68379	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
68417	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
68427	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
68498	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
68534	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
68571	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
68582	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
68629	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
68646	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
72006	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
72021	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
72055	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
72063	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
72140	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Figure 6: Arcade 3 pattern. County assignments on X-axis, Right DID on Y-axis. Number of matching records at intersection.

Icicle

The Icicle pattern, visible using the same method as Arcade, assigns records across all counties. For each Right DID range, most numbers are assigned to a single "Priority County", with smaller distributions to others. The pattern divides the full range into 21 "icicles", one per county. Priority counties contain hundreds to thousands of records per Right DID, while neighbouring counties have dozens or fewer (**Figure 7**).

Right ID	Atlantic	Bergen	Burlington	Camden	Cape May	Cumberla	Essex	Glouceste	Hudson	Hunterdo	Mercer	Middlesex	Monmout	Morris	Ocean	Passaic	Salem	Somerset	Sussex	Union	Warren
51174	1	2	8	2	0	0	521	1	5	1	0	12	7	17	16	5	0	6	3	31	3
51175	3	6	0	3	0	1	308	2	4	0	0	8	6	7	4	5	0	1	2	2	1
51176	1	3	1	5	2	0	483	2	6	0	2	7	4	16	13	2	0	2	3	8	1
51177	1	5	4	3	0	3	684	3	0	3	1	8	6	27	15	4	0	4	2	10	0
51178	2	4	4	6	3	1	876	5	3	2	3	4	13	29	15	7	0	6	1	14	3
51179	3	7	4	3	1	1	1068	3	4	0	2	8	11	28	28	15	0	9	5	20	5
51180	1	3	4	2	0	0	1211	3	3	2	4	6	9	37	28	13	0	4	3	23	5
51181	1	6	6	2	0	1	1243	3	8	1	6	16	13	34	30	16	0	8	7	19	2
51182	7	4	6	7	1	2	1326	1	3	3	9	12	19	46	39	22	0	8	3	24	1
51183	3	8	5	3	0	1	1266	0	3	2	1	15	16	37	43	14	0	8	3	35	4
51184	2	7	2	2	2	1	1217	3	6	3	9	17	17	50	47	16	1	10	2	22	3
51185	4	7	5	5	5	0	1407	0	4	4	6	12	24	32	66	29	1	8	5	36	5
51186	3	11	1	6	2	0	1320	1	3	3	7	7	17	52	59	44	0	13	3	19	5
51187	3	12	8	5	3	0	1633	3	9	3	6	26	20	42	56	34	0	10	7	41	3
51188	1	10	7	7	2	1	1679	1	7	4	3	24	27	67	75	30	0	10	12	28	2
51189	6	7	5	6	3	1	1591	3	5	4	2	18	18	67	67	26	0	12	9	29	3
51190	5	10	11	1	2	1	1637	3	4	3	3	23	22	57	76	25	1	23	7	39	5
51191	5	9	2	5	5	0	1604	4	12	4	6	24	25	61	71	27	2	18	3	33	2
51192	7	9	6	2	4	0	1795	1	9	6	4	31	24	57	61	31	0	18	4	37	9
51193	3	14	10	6	1	1	1857	4	12	10	3	30	22	62	53	21	1	8	8	44	4
51194	5	19	8	9	2	3	1737	2	8	11	6	31	20	83	39	28	0	20	5	51	6
51195	2	15	10	11	2	1	1765	3	6	7	15	17	27	94	58	24	0	19	10	62	6
51196	7	17	7	4	1	1	1707	1	11	6	5	21	27	67	52	31	0	13	6	48	5
51197	6	21	11	9	2	0	1705	1	13	8	9	26	18	81	53	31	0	20	10	49	13
51198	4	16	10	6	4	1	1560	1	20	2	5	22	22	65	33	32	0	15	9	82	9
51199	3	22	6	8	2	1	1642	3	8	19	9	30	22	66	48	23	0	15	16	42	4
51200	6	33	9	7	1	1	1637	3	10	10	8	33	25	55	46	37	0	19	15	44	9
51201	4	31	15	6	3	2	1662	0	18	8	3	44	18	58	40	27	0	17	5	91	4
51202	9	22	9	5	2	2	1644	1	29	5	9	48	29	42	30	38	0	25	11	98	7
51203	8	12	5	3	1	3	1777	0	4	1	0	20	23	58	64	18	0	11	6	45	5
51204	2	3	2	1	2	0	499	0	1	3	2	6	7	17	20	8	1	3	2	12	2

Figure 7: Icicle 1, Essex County section. Counties on X-axis, Right DID on Y-axis. Priority County highlighted in black.

The transition from Arcade 1 to Icicle 1 occurs at Right DID 51,173 (last Arcade 1) to 51,174 (first Icicle 1) (**Figure 8**). The transition region between Arcade 1 and Icicle 1, as well as other transitions, shows what appears to be some penetration of the Arcade pattern into the Icicle pattern.

While Icicle partitions contain more records than Arcade partitions, this does not suggest Icicle is simply a denser version of Arcade. Icicle 1 (2,969,091 records) far outweighs Arcade 1 (555,536 records, 18.71% of Icicle 1). The density increase in Icicle 1 is over 1000-fold on average, not the 500% that record count alone would suggest. If Icicle were merely a denser Arcade, priority counties would have 5-25 records each, not the observed 500-2,000 (**Figure 8**).

Right ID	Atlantic	Bergen	Burlington	Camden	Cape May	Cumberlai	Essex	Glouceste	Hudson	Hunterdor	Mercer	Middlesex	Monmout	Morris	Ocean	Passaic	Salem	Somerset	Sussex	Union	Warren	All
51165	0	1	1	0	0	0	0	0	0	0	2	1	0	0	0	0	0	1	0	1	0	7
51166	0	1	0	1	0	1	0	0	1	1	0	0	0	0	0	1	1	0	0	1	1	9
51167	1	0	2	0	1	0	0	2	0	0	0	1	1	0	0	0	0	1	0	1	0	10
51168	0	3	1	0	1	0	1	1	1	0	0	1	1	1	2	1	0	0	1	0	1	16
51169	0	3	0	0	0	0	0	2	1	0	1	0	0	0	0	0	0	0	0	1	0	8
51170	0	1	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	1	0	6
51171	1	0	0	0	0	0	2	0	0	0	1	1	2	0	2	1	0	1	0	0	0	11
51172	0	1	1	1	0	0	1	1	1	0	1	1	0	1	0	2	0	1	0	0	0	12
51173	2	0	0	0	0	1	0	0	2	0	0	0	0	0	1	1	0	1	2	0	1	11
51174	1	2	8	2	0	0	521	1	5	1	0	12	7	17	16	5	0	6	3	31	3	641
51175	3	6	0	3	0	1	308	2	4	0	0	8	6	7	4	5	0	1	2	2	1	363
51176	1	3	1	5	2	0	483	2	6	0	2	7	4	16	13	2	0	2	3	8	1	561
51177	1	5	4	3	0	3	684	3	0	3	1	8	6	27	15	4	0	4	2	10	0	783
51178	2	4	4	6	3	1	876	5	3	2	3	4	13	29	15	7	0	6	1	14	3	1001
51179	3	7	4	3	1	1	1068	3	4	0	2	8	11	28	28	15	0	9	5	20	5	1225
51180	1	3	4	2	0	0	1211	3	3	2	4	6	9	37	28	13	0	4	3	23	5	1361
51181	1	6	6	2	0	1	1243	3	8	1	6	16	13	34	30	16	0	8	7	19	2	1422
51182	7	4	6	7	1	2	1326	1	3	3	9	12	19	46	39	22	0	8	3	24	1	1543

Figure 8: Transition from Arcade 1 to Icicle 1

Number Deconstruction

Understanding LID and DID number algorithms requires analyzing their components: DID's Alpha and Left and Right IDs for both DID and LID. Component relationships vary with sorting methods.

Default sorting (alphabetical by names, then RegDate) and RegDate-first sorting both scramble LID and DID numbers. An LID sort groups DID Alphas alphabetically in a specific pattern (I-P, then A-P six times, ending A-G) but does not order DID numerics coherently. DID sort order doesn't correlate with LID (**Table 5**).

leg_id	Leg_ID Rank	displayId	Alpha	DID NUM Only	DID Rank
520,684,834	8	A0000153542	A	0000153542	1
150,008,080	5	A0001752815	A	0001752815	2
102,102,449	1	A0003651353	A	0003651353	3
151,233,612	6	A0003653182	A	0003653182	4
102,102,452	2	A0003951353	A	0003951353	5
112,338,959	4	A0003952068	A	0003952068	6
151,233,660	7	A0005653182	A	0005653182	7
544,804,083	9	A0006553933	A	0006553933	8
102,102,494	3	A0007651353	A	0007651353	9

Table 5: DID sort compared to LID rank show lack of rank order correlation, first nine rows, Atlantic County

Sorting by any field present in the voter rolls indicates that each field is independent of the rest, despite the partial correlation of DID Alpha ID values when records are sorted by LID. However, DID and LID numbers are not only related, but are also related so closely that some DID numbers can be accurately calculated based on LID numbers and vice versa.

When sorted by LID, DID numbers behave as if their left and right sides are swapped. Thus, DID numbers are divided into "Left DID" (leftmost five digits, 100,000-10,000,000 place values) and "Right DID" (rightmost five digits, 1-10,000 place values). Left DID cycles through all lower place values before Right DID advances, resembling a Caesar cipher shift (Luciano & Prichett 1987), This shifted arrangement is termed the Right/Left concatenation (RL Concat) of DID numbers.

LID and RL Concat values, when sorted by LID, show identical rank ordering (**Table 6**). This ranking differs from the DID Num order. Consequently, RL Concat is deemed "valid", while the DID number in voter rolls is likely an obfuscated version of RL Concat.

leg_id	DID rank	displayId	Alpha	DID NUM	DID rank	Left NUM	Right NUM	RL Concat	RL Concat Rank
101000988	1	I2812351174	I	2812351174	91,720	28,123	51,174	5,117,428,123	1
101003262	2	I2880851175	I	2880851175	93,948	28,808	51,175	5,117,528,808	2
101013639	3	I3144151176	I	3144151176	102,161	31,441	51,176	5,117,631,441	3
101024549	4	I3551651177	I	3551651177	114,723	35,516	51,177	5,117,735,516	4
101030314	5	I3797551178	I	3797551178	122,298	37,975	51,178	5,117,837,975	5
101034536	6	I3991351178	I	3991351178	128,241	39,913	51,178	5,117,839,913	6
101036381	7	I4082451179	I	4082451179	130,975	40,824	51,179	5,117,940,824	7
101037315	8	I4128351179	I	4128351179	132,393	41,283	51,179	5,117,941,283	8

Table 6: LID sort. Left ID advances before Right DID. Ranks are for full range of 201,865 Atlantic County records.

DID numbers use 16 alpha characters (A-P), mirroring hexadecimal's base-16 system. For records without LIDs, the Full Left DID range per alpha is 0-99,999. For records with LIDs, Left ID numbers map to 00000-65,535. The upper limit, 65,535, equals hexadecimal "FFFF" (2^{16}-1), the maximum four-digit hexadecimal value (**Table 7**).

Alpha	Number of Records	MIN Left ID	HEX MIN	MAX Left ID	HEX MAX	with Leg ID	MAX Left ID with Leg ID	HEX MAX Left ID with Leg ID
A	405,758	0	00000	99,888	18630	365,922	65,535	FFFF
B	386,577	0	00000	99,984	18690	343,591	65,535	FFFF
C	411,839	0	00000	99,072	18300	369,085	65,535	FFFF
D	416,043	0	00000	99,993	18699	373,152	65,535	FFFF
E	424,819	0	00000	99,968	18680	379,579	65,535	FFFF
F	426,240	0	00000	99,424	18460	385,416	65,535	FFFF
G	384,082	0	00000	97,648	17D70	336,272	65,535	FFFF
H	368,765	0	00000	99,998	1869E	320,400	65,535	FFFF
I	384,892	0	00000	99,966	1867E	336,085	65,535	FFFF
J	413,845	0	00000	99,926	18656	369,724	65,535	FFFF
K	406,233	0	00000	99,977	18689	360,379	65,535	FFFF
L	402,975	0	00000	99,760	185B0	360,083	65,535	FFFF
M	399,611	0	00000	99,976	18688	353,695	65,535	FFFF
N	417,101	0	00000	99,424	18460	366,989	65,535	FFFF
O	395,171	0	00000	99,939	18663	353,419	65,535	FFFF
P	415,482	0	00000	99,838	185FE	371,962	65,535	FFFF
	6,459,433		0	99,998		5,745,753	65,535	

Table 7*:* MIN/MAX Left ID by Alpha, full range and constrained to Legacy ID numbers only

Legacy IDs, with 16 alpha characters and a max Left ID of 65,535 (Hex FFFF), link to hexadecimal, unlike Arcade numbers. Arcade DIDs span 0-99,999, making Left ID only relevant above 65,535. This 65,535 limit, unnecessary for compression given LID and DID digits exceed population needs, oddly appears in pre-DID records. Icicle 3 uniquely increments LID by 23 when sorted by LID or RL Concat, with missing records causing multiples of 23. All Icicle patterns show DIDs incrementing by 100,000 under the same sort (**Table 8**).

LEG NUM	DID NUM	Concat RL NUM	ALPHA	LEG GAP	LEG GAP/23	DID Gap	Concat Gap
520,002,539	3,587,353,533	5,353,335,873	P	23	1	100,000	1
520,002,562	3,587,453,533	5,353,335,874	P	23	1	100,000	1
520,002,585	3,587,553,533	5,353,335,875	P	23	1	100,000	1
520,002,608	3,587,653,533	5,353,335,876	P	23	1	100,000	1
520,002,631	3,587,753,533	5,353,335,877	P	23	1	100,000	1
520,002,654	3,587,853,533	5,353,335,878	P	23	1	100,000	1
520,002,677	3,587,953,533	5,353,335,879	P	23	1	100,000	1
520,002,700	3,588,053,533	5,353,335,880	P	23	1	100,000	1
520,002,723	3,588,153,533	5,353,335,881	P	23	1	100,000	1
520,002,746	3,588,253,533	5,353,335,882	P	23	1	100,000	1
520,002,769	3,588,353,533	5,353,335,883	P	23	1	100,000	1

Table 8: Icicle 3 LID numbers increment by 23 when sorted by RL Concat

Priority counties

New Jersey's three Icicle patterns, when sorted by RL Concat, present the state's 21 priority counties non-alphabetically (**Table 9**). Essex has the lowest Right DID values, followed by Gloucester, Mercer, Middlesex, and Ocean. This resembles New York's Spiral algorithm, where 67 sub-partitions exist: 62 for counties and 5 for unassigned State Board of Election Identification (SBOEID) numbers. Sorting by SBOEID disrupts both the alphabetical order of New York's County Codes and the alphabetical Federal Information Processing Series (FIPS) codes.

Atlantic	Bergen	Burlingtor	Camden	Cape May	Cumberla	Essex	Glouceste	Hudson	Hunterdo	Mercer	Middlesex	Monmout	Morris	Ocean	Passaic	Salem	Somerset	Sussex	Union	Warren
						1	2			3	4			5					6	
7				8	9			10								11				
			12						13			14						15		
													16				17			18
	19	20													21					

Table 9: Arrangement of county rank priority based on Right DID

In Icicle 1, Left LID numbers uniquely encode county information. The second and third digits indicate both county name and order within the Icicle pattern (**Table 10**). Each Left LID is exclusive to its originating county. The digits (101, 102, 103, for example) reveal the county's order when sorted by RL Concat. This county-encoding feature is absent in Icicle 2 and Icicle 3, making it a hidden attribute exclusive to Icicle 1 records.

County Name	County number	County FIPS code	Order	Second and third num Leg ID	Lowest Leg ID	Highest Leg ID
Essex	7	34013	1	01	101,000,108	101,668,185
Gloucester	8	34015	2	02	102,000,006	102,191,462
Mercer	11	34021	3	03	103,000,005	103,360,051
Middlesex	12	34023	4	04	104,300,005	104,711,725
Ocean	15	34029	5	05	105,001,004	105,704,950
Union	20	34039	6	06	106,000,003	106,370,850
Atlantic	1	34001	7	07	107,000,045	107,198,851
Cape May	5	34009	8	08	108,000,002	108,069,451

Table 10: Relationship of Left LID to county when sorted by Right DID and then Left DID

The three Icicle patterns possess distinct traits that serve as partition tags. Icicle 1 uniquely encodes county and order in the LID's second and third digits. Icicle 3 distinctively increments LID numbers by 23. Icicle 2 lacks both these features, distinguishing it from Icicles 1 and 3.

New York Algorithms

New York employs four algorithms: Spiral, Metronome, Tartan, and Shingle, distributed across three primary partitions: Out of Range Low (ORL), In-Range (IR), and Out of Range High (ORH). Spiral and Metronome occupy only IR, Shingle only ORH, while Tartan spans both ORL and ORH. This structure parallels New Jersey's two algorithms across six partitions.

The IR partition contains 62 sub-partitions of contiguous numbers for each New York county. In contrast, ORL and ORH fragment county ID numbers, creating noise that conceals the more orderly IR numbers. Similarly, in New Jersey, orderly Icicle records are flanked by pseudo-random noise from the Arcade algorithm (**Table 11**).

DOB	RegDate	county	Leg_ID	displayID	Alpha ID	Left 5 ID NUM	Right 5 ID NUM
08/30/2002	09/17/2020	Middlesex		E2426416147	E	24264	16147
11/20/1960	10/06/2020	Ocean		M2977416147	M	29774	16147
04/04/1973	09/21/2020	Bergen		N5098016147	N	50980	16147
04/20/1941	09/02/2021	Ocean		A5173816147	A	51738	16147
09/04/2000	06/05/2020	Camden		F6550616147	F	65506	16147
12/14/1994	09/17/2020	Warren		M0125016148	M	1250	16148
06/22/1962	03/23/2021	Ocean		D2499816148	D	24998	16148
03/08/1991	06/07/2021	Middlesex		O3922516148	O	39225	16148
11/04/1988	11/14/2020	Hudson		P4261216148	P	42612	16148

Table 11: Arcade records sorted by RL concat have disordered alpha and county values

New Jersey and New York ID number algorithms share eight key characteristics:

1. Partitioned state space
2. Central ordered sub-partitions flanked by pseudo-random ends
3. One sub-partition containing all county records (NJ: Icicle 1)
4. Non-alphabetical county order
5. Distinctive incremental spacing in one partition (NJ: Icicle 3)
6. Algorithmic obfuscation of ID numbers (NJ: Icicle 1-3)
7. Shift cipher-like manipulation
8. Covert record-referencing capability through algorithmic manipulations

Discussion

New Jersey's voter rolls employ two ID types: DID (universal) and LID (applied to ~88% of records). Both exhibit complexity beyond simple serial numbering.

LID numbers can be used for multiple purposes:

1. Differentiating Icicle from Arcade patterns
2. In Icicle 1, identifying county of origin and its order within the pattern
3. Through incremental spacing, distinguishing Icicle 3 from Icicles 1 and 2

DID numbers, while appearing as serial numbers, reveal hidden relationships when transformed:

1. Shifting the first five digits right and wrapping the last five to the front
2. Sorting by this transformed value matches LID sort order exactly

3. This transformation obfuscates the LID-DID relationship

The Alpha component of DID numbers shows no apparent meaning across the full Left DID range. However, within Icicle partitions:

1. Left DID numbers for each alpha value are limited to 0-65,535 (hex FFFF)
2. This limitation appears artificial and unrelated to LID structure
3. Left DID numbers above 65,535 can be used to accurately identify Arcade records

The DID and LID number treatments enable accurate identification of two algorithms across five partitions. While the purpose of these differences beyond partition and algorithm segregation is unknown, they could potentially facilitate covert ID tracking if specific algorithms were assigned to targeted records.

Large datasets often segregate records for legitimate reasons, such as distinguishing healthcare insurance customers by age, priority, insurance type, geography, or medical condition. Typically, this information is stored in dedicated database fields for easy access. Some identification systems, like Social Security Numbers, encode geographic data within the number itself, allowing location determination with appropriate reference tables.

New Jersey's Icicle 1 LID numbers embed county of origin information, a feature absent in Icicles 2 and 3. This county information is combined with a non-alphabetical county order, similar to New York's system. The rationale for including this information in only one of six partitions, and the advantage of the scrambled order, remain unclear.

Including county of origin in identification numbers is a common practice. Pennsylvania, for example, appends two-digit county codes to voter ID numbers (for example, "001018014-01"), using a hyphen to distinguish the county code from the unique serial ID. These county codes are in alphabetical and ascending numeric order for ease of use. In contrast, New Jersey's approach of embedding this information only in Icicle 1 LID numbers, without clear delineation and in a non-alphabetical order, appears less transparent and potentially less useful.

Data masking is widely employed for legitimate purposes across various fields. It's used to enhance security in payment processing (Hand 2010), to protect privacy in phone calls (Becker *et al.* 2010), to maintain patient confidentiality in medical studies (Goroff *et al.* 2018), and to secure military communications (Kahn1980). Common masking techniques include using partial identifiers, such as the last four digits of Social Security or credit card numbers, to prevent unauthorized disclosure and use while still allowing necessary identification.

The shifting of components in New Jersey's DID numbers represents an obfuscation technique typically used to protect sensitive information. Paradoxically, the obfuscated data remains publicly accessible, raising questions about the purpose of this complexity in a public dataset.

Both New Jersey and New York employ multi-layered obfuscation algorithms in their voter ID systems. In New York, as detailed earlier, these algorithms allow for unexpected capabilities, such as predicting voter status through the Shingle algorithm and identifying deleted or cloned records via the Spiral algorithm. The ability to identify cloned records is particularly concerning, as such records are illegal. Moreover, the link between ID numbers and voter status or record type deviates from expected data management practices.

New Jersey's voter rolls lack sufficient detail to verify if its algorithms can predict attributes similar to New York's Shingle and Spiral algorithms. However, organizational similarities between the states' systems suggest potential parallels. Both states, for instance, contain cloned records, which are illegal.

While New Jersey's known clone count (18,067) is significantly lower than New York's approximately 2 million, many of New York's clones are in purged or deleted records. This might seem to negate fraud concerns, as purged records theoretically cannot be used to vote. However, 1,488,251 of New York's "purge" status records lack purge dates, making it impossible to confirm their status during past elections. Tens of thousands of these records show votes potentially cast post-purge, although this cannot be verified definitively (NYCA 2022). Additionally, 254,713 votes recorded in New York City's County rolls are absent from state records. New Jersey's data is insufficient to check for similar discrepancies.

Potential benign explanations for the observed patterns in New Jersey's voter ID system have been considered. However, given the public nature of the database, the complexity of the algorithms, their inconsistent application, and the creation of 'shadow data', no plausible innocent justification was found. Standard practices in data management, including those for public records, do not account for these characteristics. The presence of illegal clone records further undermines potential benign explanations. The deliberate obfuscation in a system legally required to be transparent raises significant concerns. This leads to the conclusion that these algorithmic manipulations likely serve a purpose beyond any legitimate data management need, warranting serious scrutiny and further investigation.

Conclusion

New Jersey's DID numbers show evidence of algorithmic modification, suggesting potential obfuscation. While the intent behind these alterations is unclear, their presence raises questions about compliance with public disclosure laws. Election law mandates transparency for auditability and public confidence. The use of modified ID numbers could be perceived as concealment, potentially conflicting with legal requirements and providing opportunities for data obscurity.

1. Unnecessary Complexity: The voter ID system is far more complex than required for its ostensible purpose. A simple sequential numbering system would suffice for unique identification.
2. Data Obfuscation: The use of shift cipher-like techniques and complex algorithms to generate and manipulate ID numbers is a red flag. In a public database, this level of obfuscation is unnecessary and counterproductive to transparency.
3. Partitioned Data: The division of data into distinct partitions (Arcade and Icicle) with different characteristics suggests a deliberate attempt to categorize or segregate records in a non-obvious way.
4. Hidden Attributes: The ability to derive information (like county of origin) from certain ID numbers, but only in specific partitions, indicates the presence of embedded data not visible in standard fields.
5. Inconsistent Application: The fact that different algorithms are applied to different subsets

of the data is suspicious. It suggests that records are being treated differently for an undisclosed purpose.
6. Potential for Covert Tracking: The complex relationship between LID and DID numbers, and the ability to transform one into the other, could be used to track or categorize records without obvious flags.
7. Overengineering: The system appears vastly overengineered for its stated purpose, which is typically a sign that it's being used for undisclosed functions.

This pattern would strongly suggest a system designed for covert data management or manipulation. The complexity and obfuscation techniques are consistent with methods used to hide certain data or operations from casual observation while maintaining the ability to access or manipulate records in specific ways if one knows the system.

Given that this is a public voter database, which should prioritize transparency and simplicity, these characteristics are extremely concerning. They suggest either severe mismanagement or, more worryingly, a deliberate attempt to create a system that can be manipulated in ways not apparent to outside observers. This kind of system architecture is more commonly associated with covert operations or systems designed to resist auditing and oversight.

The worst-case scenario suggested by this study's findings is that malicious actors have compromised New Jersey's voter rolls for the purpose of committing election fraud. The best-case scenario is that an over-engineered group of algorithms was utilized, for no recognizable purpose, to assign ID numbers in a manner that unnecessarily complicates database management and introduces the risk of error and higher cost of maintenance.

References

Balmer, B 2017, 'Shadow values and the politics of extrapolation', *Science, Technology, & Human Values*, vol. 42, pp. 311-14.

Baptist, J & Gluck, J 2021, 'The gray legion Information Warfare within our gates', *Journal of Strategic Security*, vol. 14, pp. 37-55.

Becker, RA, Volinsky, C & Wilks, AR 2010, 'Fraud detection in telecommunications: History and lessons learned, *Technometrics*, vol. 52, pp. 20-33.

Downes, C 2018, 'Strategic blind-spots on cyber threats, vectors and campaigns', *The Cyber Defense Review*, vol. 3, pp. 79-104.

Ex Parte Coy 1888, *US 731*, Supreme Court of the United States, Washington, D.C., US.

Goroff, D, Polonetsky, J & Tene, O 2018, 'Privacy protective research facilitating ethically responsible access to administrative data', *The Annals of the American Academy of Political and Social Science*, vol. 675, pp. 46-66.

Hand, DJ 2010, 'Fraud detection in telecommunications and banking: Discussion of Becker, Volinsky, and Wilks (2010) and Sudjianto *et al.* (2010)', *Technometrics*, vol. 52, pp. 34-8.

Help America Vote Act 2002, *Public Law 107-252*, US Congress, Washington, D.C., US.

Hodgson, QE, Brauner, MK & Chan, EW 2020, 'Securing US Elections against cyber threats; Considerations for supply chain risk management', RAND Corporation, viewed 4 July 2023, <https://www.rand.org/pubs/perspectives/PEA512-1.html>.

Hornik, M & Paquette, A 2022. 'New York's 2020 general election: A study in deficits', New York Citizen's Audit.

Kahn, D 1980, 'Codebreaking in World Wars I and II: The major successes and failures, their causes and their effects', *The Historical Journal*, vol. 23, pp. 617-39.

Kanuck, S 2022, 'Content as infrastructure: The unlearned lesson about cyber security an information integrity', *The Cyber Defense Review*, vol. 7, pp. 181-92.

Luciano, D & Prichett, G 1987, 'Cryptology: From Caesar ciphers to public-key cryptosystems', *The College Mathematics Journal*, vol. 18, pp. 2-17.

National Voter Registration Act 1993, *103-31*, House of Representatives, Washington, D.C., US.

Paquette, A 2023, 'The Caesar cipher and stacking the deck in New York State voter rolls', *Journal of Information Warfare*, vol. 22, no. 2 , pp. 86-105.

Templ, M & Sariyar, M 2022, 'A systematic overview on methods to protect sensitive data provided for various analyses', *International Journal of Information Security*, vol. 21, pp. 1233-46.