Wisconsin voter roll ID numbers

Preliminary Report

Andrew Paquette

10/01/2024

Introduction

The investigation of Wisconsin voter rolls began in New York.

New York State's voter rolls contain many irregularities. Among them, an estimated 2,000,000 illegal "clone" records. A clone is a record with two or more unique state identifiers, or State ID (SID). Clones are illegal under state and federal law because they allow for the generation of multiple ballots per voter per election.

If the clones were generated by administrative error or other innocent cause, there should be no way to track them separately from other records. However, if created for malicious use, malicious actors would want a way to covertly find the records. This led me to investigate whether voter roll data fields contained hidden attributes for covert tracking. The SID and County ID (CID) fields were the most likely candidates, given their uniqueness compared to names, birthdates, addresses, and other fields.

The investigation revealed that at least 4 algorithms were used as part of the ID assignment process (Paquette 2023). One algorithm allows very accurate predictions (99.34%) of voter status and whether the record is a clone. Another algorithm reveals deleted SID numbers and who those numbers were originally assigned to. This second algorithm, called "The Spiral," also adds an attribute to all records that is effectively a third, and very well-hidden, ID number. An Algorithm ID (AID).

The results of the NY investigation led me to check the records contained in New Jersey's voter rolls (Paquette 2024). These also revealed a pseudo-cryptographic artifact in their ID numbering system. It transformed the numbers from an original state to what is visible in the rolls. The transformation is reversible, making it possible to use the structure of the transformation as a way to covertly identify records of interest.

An investigation of Ohio's voter rolls revealed hidden attributes attached to records in three of Ohio's most populous counties: Franklin, Lucas, and Montgomery. Those attributes would allow for covert tracking and identification of records of interest.

Separately, researcher Vico Bertogli of Pennsylvania discovered that Hawaii's voter rolls contain a tagging mechanism on their ID numbers. Hawaii uses a 32-digit encrypted ID called a "UUID". Slightly less than ten percent of all records have the same final 12 digits. All other records do not have this trait. As in other states, this difference segregates one group of records from the rest.

A fundamental rule of database management is that all data should be transparent, traceable, and used only for its intended purpose. The algorithms found in various state databases violate this rule by introducing what amounts to undocumented attributes into the database. This makes it untraceable by normal means and can enable manipulations that violate the intended purpose of the databases.

I was approached by a concerned citizen of Wisconsin who was aware of my research in other states to review Wisconsin's voter rolls.

This report is preliminary, aiming to:

1. Quickly determine if hidden attributes have been added to Wisconsin's voter roll records.
2. Assess whether these attributes were added by deterministic (predictable) algorithm(s) without county or state officials' knowledge.

The focus is not on whether "an algorithm was used" - all ID assignment software uses algorithms. Instead, we're interested in algorithms that:

- Are over-engineered for ID assignment
- Employ obfuscation techniques
- Add hidden attributes allowing data segregation

The presence of algorithms can be determined without understanding what they do. Their purpose can be determined without solving them. Solving an algorithm makes it completely reversible, but this step isn't necessary to demonstrate the presence or purpose of an algorithm, or even that it is reversible. In NY, it took about a week to prove the presence of multiple algorithms. It took about a month to see approximately what they were doing, but it took two years to solve the first of four.

Initial results show that Wisconsin has a significant number of illegal cloned records in their most current database (August, 2024). These exist in sufficient numbers (n=874,455) to make usage of ID-tracking algorithms worthwhile. Many counties reflect the use of multiple algorithms to assign ID numbers. Broadly, they fall into the following 3 categories:

- Increment by one (normal)
- Increment by variable multiple of ten
- Increment method designed to appear random

Preliminary results

Scatterplots

For the first stage of this analysis, Wisconsin's voter ID numbers were compared to registration dates. The analysis showed that many ID numbers are not perfectly assigned in chronological order, though some are.

In a scatter plot, sequential assignment creates a sloped line starting at the lower left corner of the graph (the lowest ID numbers) and ascend toward the upper right (highest numbers). Many of Ohio's 88 counties have this pattern, such as Fairfield County (Figure 1).

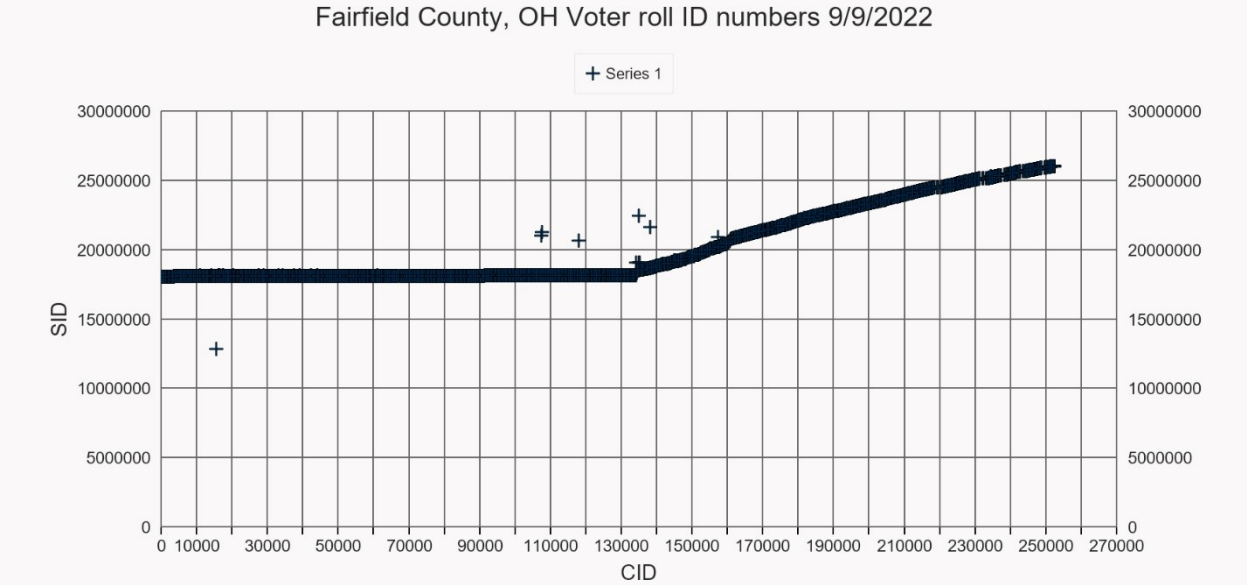


Figure 1 Fairfield County, Ohio, sequential CID and SID numbers

Wisconsin's scatterplots, like Milwaukee County's (Figure 2), show unusual patterns. They display segregated number blocks, suggesting varied ID assignment methods. Overlapping date ranges within each block indicate simultaneous use of multiple approaches, pointing to complex factors or algorithmic controls.

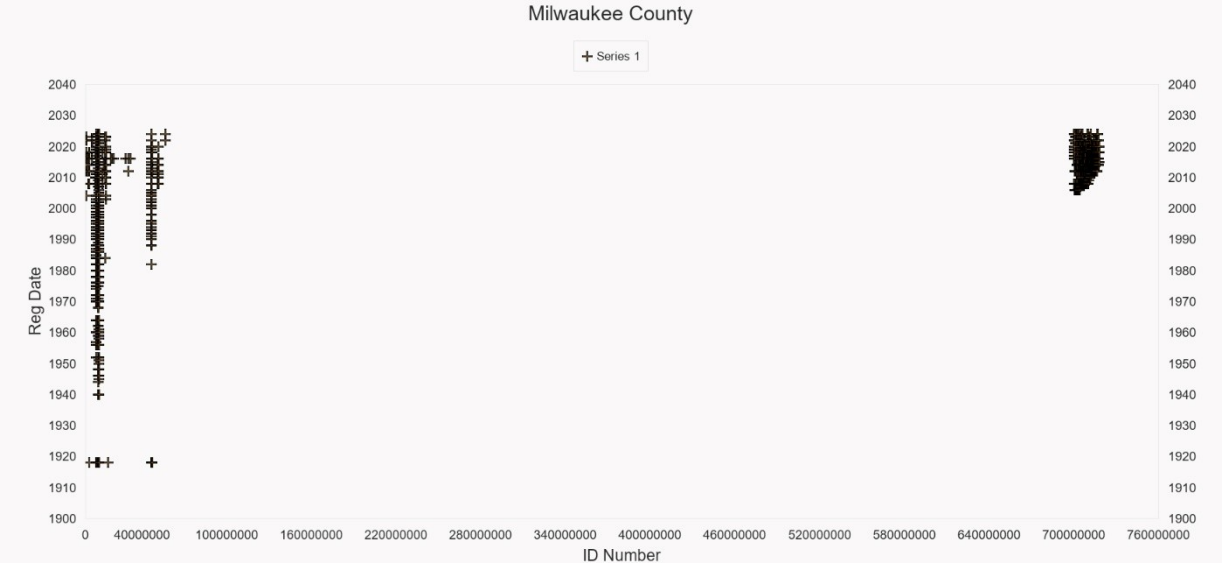


Figure 2 Milwaukee County ID numbers by Registration Date

A closeup of the Wood County scatterplot reveals details obscured in the full-range view (Figure 3). ID numbers are not assigned strictly sequentially by date. The plot shows a clear shift in ID number assignment around 2016. From 2006 to 2015, there's a pattern of ascending ID numbers. In 2016, there's an abrupt change, with new registrations receiving ID numbers similar to those used a decade earlier. This pattern continues through 2024, suggesting a significant reset or change in the ID assignment system in 2016.

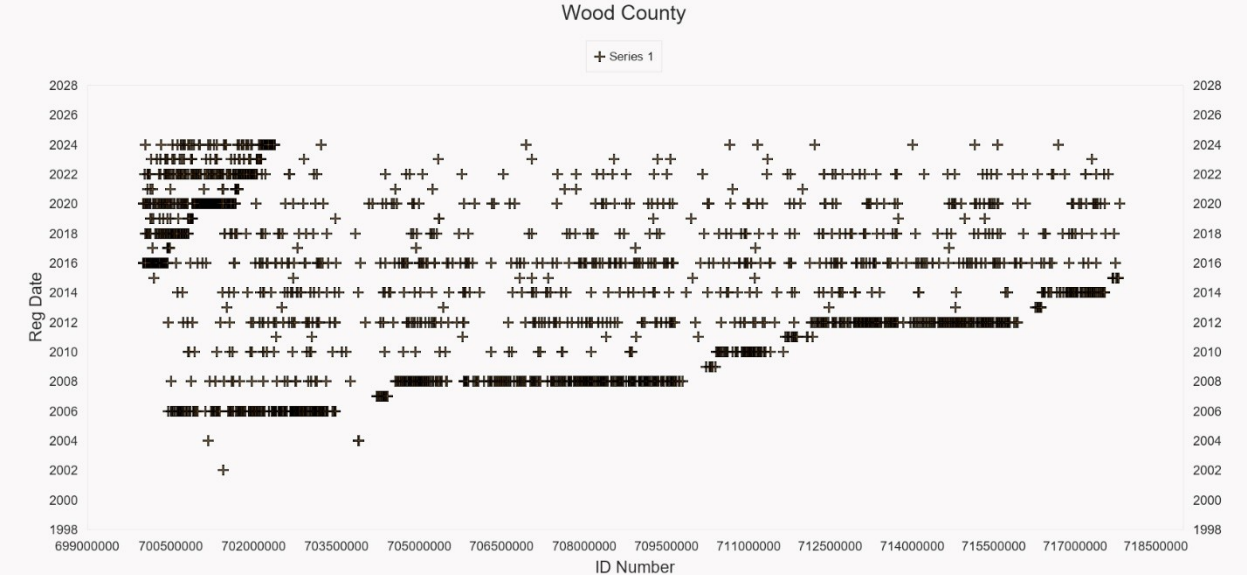


Figure 3 Wood County ID numbers, close-up

Wisconsin's scatterplots do not resemble the "normal" Fairfield County plot. It is possible that the structure evident in Wisconsin's ID number plots can be explained as any of the following (or something else unknown):

- Numbers assigned in different eras to differing assigned ranges, combined with movement of voters within the state, linking newer registrations to older ID numbers.
- The structures evident in Wisconsin could segregate records based on attributes. If so, and those attributes aren't contained in data fields as they should be, this would qualify as hidden attributes. Hidden attributes violate normal database administration protocols and are generally considered unethical.
- Unused white space may be reserved space for future registration. If sequential numbering is used, there is no need to reserve numbers like this because there is an infinity of available numbers. Only if the structure can go backwards by assigning smaller numbers after larger numbers, is this kind of reservation technique required. If that is what is done here, it is non-standard and should be explained, particularly given the large number space used in Wisconsin counties.
- System evolution. This would mean that the database designers have changed the methods used to assign ID numbers several times. If standard numbering were used, there would have been no need. Standard numbering offers multiple advantages: ease of implementation, transparency, usability, cost-effectiveness, and reduced administrative overhead.

Until more is known, it is difficult to say why these ID numbers are organized as they are. However, they are unusual enough to warrant further scrutiny.

Gap analysis

Gap analysis compares two values by subtracting one from the other to determine the difference. In New York, this method revealed a 'Spiral' algorithm using Repunit-based patterns (e.g., 1,111, 111, 11) in

SID numbers. This pattern was consistent, with predictable variations for missing records. Gap analysis can be affected by added or deleted records, changing gap values unpredictably.

New York's gap distribution was highly organized, following a predictable pattern: every 10th gap was 11, every 100th was 111, every 1000th was 1,111, and so on. Other predictable values were interspersed.

In Wisconsin, this analysis did not reveal any clear distribution patterns. However, certain gap values recurred at higher than expected rates, suggesting a non-random assignment process.

Gap sizes in a natural distribution are inherently biased by the mathematics of subtraction. Smaller gaps occur more frequently because they can be produced by a wider range of number combinations. Conversely, larger gaps are rarer, as fewer number pairs can produce them when subtracted. This relationship between gap size and frequency is illustrated in Table 1, showing an inverse correlation between gap magnitude and its likelihood of occurrence.

Table 1 Subtraction product distribution based on number pair combinations

0	1	2	3	4	5	6	7	Gap	Total combos	Pct
1	0	NA	NA	NA	NA	NA	NA	0	7	25.00%
2	1	0	NA	NA	NA	NA	NA	1	6	21.43%
3	2	1	0	NA	NA	NA	NA	2	5	17.86%
4	3	2	1	0	NA	NA	NA	3	4	14.29%
5	4	3	2	1	0	NA	NA	4	3	10.71%
6	5	4	3	2	1	0	NA	5	2	7.14%
7	6	5	4	3	2	1	0	6	1	3.57%

The general expectation is that in a random distribution of numbers, gap frequencies would follow this natural pattern. Most Ohio counties demonstrate this expected descending gap frequency distribution (Table 4), with only three exceptions (two illustrated-Lucas and Montgomery Counties).

Table 2 Sample Ohio counties show descending gap frequencies, with 2 exceptions in Lucas and Montgomery Counties

CID Gap Value	Lawrence Frequency	Licking Frequency	Logan Frequency	Lorain Frequency	Lucas Frequency	Madison Frequency	Mahoning Frequency	Marion Frequency	Medina Frequency	Meigs Frequency	Mercer Frequency	Miami Frequency	Monroe Frequency	Montgomery Frequency
1	24,297	68,766	17,901	126,786	121,830	13,898	81,955	21,284	74,315	8,224	17,125	40,114	5,105	149,615
2	7,174	22,141	6,210	39,467	35,113	4,362	28,763	8,167	22,661	2,846	5,489	13,296	1,794	47,989
3	3,856	11,651	3,246	19,672	15,591	2,370	15,548	4,135	11,728	1,570	2,811	7,223	1,002	23,378
4	2,370	6,899	1,805	11,137	7,691	1,480	9,711	2,450	7,005	930	1,539	4,227	592	12,150
5	1,616	4,471	1,096	6,699	4,261	1,015	6,486	1,528	4,382	599	915	2,908	341	6,870
6	1,146	2,787	684	4,209	2,388	705	4,447	912	2,897	370	559	1,951	248	4,015
7	815	1,921	447	2,761	1,476	500	3,341	562	1,962	247	330	1,438	180	2,511
8	629	1,374	257	1,924	28,301	366	2,531	417	1,378	176	220	953	83	20,905
9	464	961	184	1,360	526	274	1,891	260	954	117	125	704	82	1,403
10	377	707	100	939	320	199	1,570	158	706	97	77	527	41	601
11	282	483	59	696	299	142	1,312	144	517	54	54	395	29	660
12	217	368	42	539	144	122	1,071	78	384	53	32	315	20	366
13	187	258	35	411	94	86	859	41	273	31	16	277	12	263

In contrast, Dane County, Wisconsin, like other Wisconsin counties, exhibits a gap frequency pattern that deviates significantly from this expected distribution. This deviation is influenced by the use of three distinct algorithms to assign ID numbers (Table 3).

Table 3 Dane County, WI gap frequency types

Gap Type	MIN	RegDate	MAX	RegDate	Range	Record count	Pct range used
Consecutive	40,002	1/1/1918	990,279	11/2/2004	950,278	384,285	40.44%
Randomized	1,650,094	09/11/2024	702,371,662	09/16/2024	700,721,569	384,591	0.05%
10 Multiples	702,371,720	11/07/2006	717,827,930	8/26/2020	15,456,211	179,758	1.16%

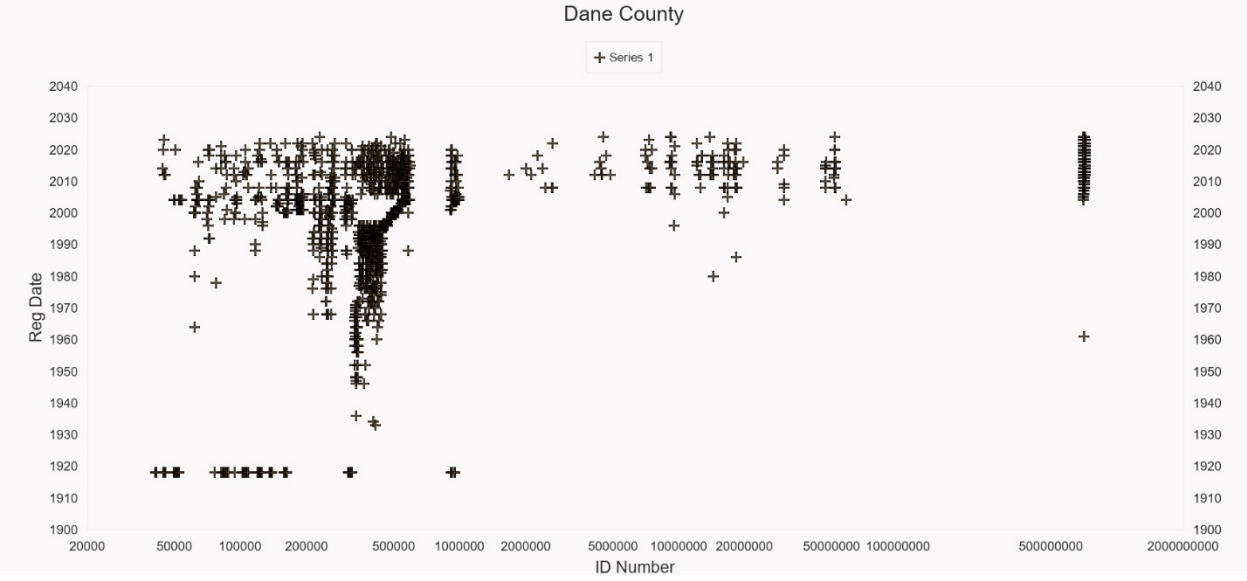
The most prominent effect is the prevalence of gaps that are multiples of 10. This is due to the third algorithm (associated with the highest ID numbers) incrementing by multiples of 10 (Table 4). The order of these numbers appears randomized, leading to sequences of gaps like so: 20, 120, 50, 120, 430, 190, 50, 480, 40, 570, 10.

Table 4 Dane County, WI, first 50 gap frequencies compared

GAP	FREQ	GAP	FREQ	GAP	FREQ	GAP	FREQ	GAP	FREQ
1	95,716	11	7,925	21	1,977	31	717	41	337
2	48,404	12	6,647	22	1,725	32	615	42	344
3	35,366	13	5,720	23	1,500	33	599	43	312
4	28,359	14	5,072	24	1,391	34	570	44	288
5	22,426	15	4,252	25	1,256	35	528	45	274
6	18,652	16	3,863	26	1,151	36	475	46	288
7	15,875	17	3,305	27	1,028	37	456	47	202
8	13,140	18	2,871	28	913	38	410	48	269
9	11,124	19	2,688	29	864	39	417	49	245
10	**65,898**	**20**	**25,907**	**30**	**15,389**	**40**	**11,331**	**50**	**8,834**

Dane County's ID numbers are segregated into easily identified clusters (Table 5). Notably, these segregated ID numbers significantly overlap between the years 2004-2024. This overlap creates the possibility that the different number ranges utilized encode an obfuscated attribute - potentially a way of categorizing voters that is not explicitly stated in the public records. Such hidden categorization, if present, could raise concerns about transparency and equal treatment in the voter registration process.

Table 5 Dane County, Wisconsin ID number layout



This unusual pattern in Wisconsin, particularly when compared to the more typical distribution seen in most Ohio counties, underscores the need for further investigation into the voter ID assignment process in Wisconsin to ensure fairness and transparency in the state's election systems.

Suspicious records

Clones

The Wisconsin voter rolls I examined lacked a Date of Birth (DOB) field; one of three key identifiers used by election clerks to check for existing voter IDs when processing new registrations. The other two are first and last names. In New York, these three fields were sufficient to identify approximately 1.5 million redundant voter ID numbers.

Generation of excess ID numbers is illegal in all states, as it potentially enables multiple votes by a single person. Each unique ID number is treated as a distinct individual by election boards. State and federal laws mandate coordination between county and state Boards of Elections (BOEs) to prevent 'cloned' records - those sharing all personal identifying information (PII) but with unique ID numbers. This differs from 'duplicates,' which may have the same ID number but vary in other details like address.

Despite lacking DOB information, Wisconsin's voter rolls include phone number and email address fields, which are more unique identifiers than DOB. I created two match fields: 'FirstName & LastName & Phone Number' (Match Name/Phone) and 'FirstName & LastName & Email' (Match Name/Email).

A Match Name/Phone search yielded 2,471,519 records out of 7,744,986. After removing records without phone numbers, 559,318 remained. Of these, 558,087 had unique ID numbers. Each 'clone' is based on an original record, with some originals having multiple clones. This indicates a minimum of 279,043 excess voter ID numbers, each potentially violating federal election law.

A similar Name/Email search, excluding empty email fields, identified 298,075 records. Combining these with the Name/Phone results yielded 874,455 suspicious records, with at least 437,227 likely clones. The use of email and phone numbers as identifiers minimizes the likelihood of false matches, as it's improbable for two people with the same name to share these contact details.

The remaining two million records likely contain additional clones, as visual inspection reveals apparent clones on most screens. For example, women changing surnames after marriage may have identical first names, addresses, and email addresses, but different last names. While this group was not included in the count due to potential false positives, it likely contains a substantial number of additional clones.

While I haven't discussed these findings with Wisconsin officials, I have conferred with multiple county Board of Elections (BOE) commissioners in New York about similar findings. Two commissioners explicitly acknowledged that the records I presented were either real and existed in their database, or were similar to records they knew of in their databases. They both told me that clone records violate state and federal law.

One commissioner explained that preventing cloned records was beyond his control due to multiple legal sources for registration applications, claiming it was impossible to prevent simultaneous processing of forms for the same voter. However, this explanation fails to account for cases where multiple ID numbers had photographic reproductions of the same signature, indicating a single origin rather than disparate sources.

The number of clone records varies considerably from county to county. The lowest percentage is in Forest County, where 2.04% of all records are identified as belonging to a clone group. The highest is 10.01% of all records, in Eau Claire. Other notable counties are: Winnebago (9.60%), LaCrosse (8.88%), Brown (8.86%), Waukesha (8.76%) Dane (8.26%), and Milwaukee (8.04%).

Of greater concern is that the number of new clone records generated per year has been rapidly increasing since the introduction of the Help America Vote Act (HAVA) in 2002. Until 2002, the number of new clone registrations never exceeded 3.00% of the total registrations in any given year (Table 6)

Table 6 Clone registrations by year, Wisconsin

WISCONSIN	<1990	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	TOTAL
Clones	29,380	732	263	4,070	272	1,180	490	3,213	294	2,143	555	7,297	821	3,494	862	23,230	4,747	17,037	100,080
Total	1,023,677	31,207	9,075	144,409	9,988	49,030	16,086	128,586	10,788	85,525	18,729	257,230	27,711	117,091	25,279	604,723	121,669	336,608	3,017,411
Pct	2.87%	2.35%	2.90%	2.82%	2.72%	2.41%	3.05%	2.50%	2.73%	2.51%	2.96%	2.84%	2.96%	2.98%	3.41%	3.84%	3.90%	5.06%	3.32%
	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	
Clones	1,704	32,707	1,063	11,395	4,716	47,888	1,399	28,968	3,061	94,267	11,058	101,338	25,675	213,420	22,969	86,860	25,894	58,977	773,359
Total	26,112	499,181	15,903	153,017	57,782	500,376	16,763	241,991	27,860	671,294	44,564	504,077	94,602	912,333	64,131	487,668	114,103	293,994	4,725,751
Pct	6.53%	6.55%	6.68%	7.45%	8.16%	9.57%	8.35%	11.97%	10.99%	14.04%	24.81%	20.10%	27.14%	23.39%	35.82%	17.81%	22.69%	20.06%	16.36%

In 2003, for the first time, clones crossed the 3.00% barrier to 3.41%. From then on, the number of clones increases year over year, with spikes in presidential election years. The highest number recorded to date is 35.82% of all registrations in 2021, and currently rests at 20.06% for the still incomplete year 2024. This radical increase in the number of plainly illegitimate registrations implies that electronic record-keeping systems have created a problem much larger than any problem they (and HAVA) intended to solve.

Encrypted ID numbers

A subset of 61,580 Milwaukee County records contained encrypted ID numbers in the format "uqMVbwqklk", using 64 characters typical of 64-bit encryption. Each encrypted ID corresponds to an unencrypted ID in the format 700,882,113 in the full database. Of these encrypted ID numbers, 53,599 (86.97%) are connected to absentee ballots.

Milwaukee County has 1,492,309 records total. It's unclear if all have corresponding encrypted IDs or if other counties use similar encryption.

The purpose of these encrypted IDs is questionable:

1. If intended for privacy or security, it's ineffective as voter information is publicly available by law.
2. Encryption in a public database is unnecessary for normal legitimate purposes.
3. This raises concerns about potentially concealing undocumented attribute information.

The extent and purpose of this encryption across Wisconsin's voter rolls require further investigation.

Comments

This study of Wisconsin's voter rolls reveals strong evidence of multiple ID number assignment algorithms, with two appearing over-engineered, enabling data segregation and hidden attribute assignment. The presence of over 60,000 records with encrypted ID numbers further suggests hidden attributes, potentially violating public disclosure laws.

Wisconsin's clone records significantly outnumber those in other states, potentially matching New York's total despite having only a third as many registered voters. This far exceeds any normal error rate or acceptable administrative standard. Alarmingly, new clone records have increased dramatically year-over-year for over two decades, potentially surpassing legal registrations within years.

These findings indicate potentially unethical management of Wisconsin's voter roll records. Regardless of intent, the algorithm's use creates a hidden classification system for data segregation, posing a security risk. The large number of cloned records exacerbates this risk, as such records would be of particular interest to those seeking to misuse voter rolls - a concern recently realized when Wisconsin mailed absentee ballots to inactive voters.

Wisconsin should investigate the algorithm's implementation in their database, specifically:

- When it was introduced
- Who introduced it
- Its intended purpose
- Associated costs
- Prior awareness among officials

Wisconsin should also consider full deletion of all clone records, particularly those with inactive status. There can be no legitimate purpose satisfied by the retention of illegal records. If there is a concern about losing voter history data, the records can be merged so that only one ID number remains per voter, as required under federal law.

This investigation is crucial for maintaining electoral integrity, data security, and public trust in the democratic process.

References

Paquette, A. (2023). "The Caesar cipher and stacking the deck in New York State voter rolls " Journal of Information Warfare **22**(2): 86-105.

Paquette, A. (2024). "New Jersey voter ID numbers reconfigured with shift cipher." (In-Press).