Supplementary Material for: Zero-Downtime Hardening of Legacy Patient Identifiers Against Transcription Errors: A Table-Free Verhoeff Formulation

Csaba Balogh

Online Resource 1: Worked Example

Let the TAJ be $\mathbf{t} = (0, 7, 6, 5, 4, 3, 2, 1, 0)$ in right-to-left order (i.e. the printed form is 012345670). We trace the TAJ-FC computation step by step.

i	t_i	$\pi_{i+1}(t_i)$	S after $\mathrm{Mul}(S, \varphi(\pi_{i+1}(t_i)))$	Note
0	0	1	$(1, 0)$	$\varphi(1) = (1, 0)$
1	7	5	$(1, 1)$	$\varphi(5) = (0, 1)$
2	6	5	$(1, 0)$	$\varphi(5) = (0, 1)$
3	5	3	$(4, 0)$	$\varphi(3) = (3, 0)$
4	4	7	$(1, 1)$	$\varphi(7) = (2, 1)$
5	3	5	$(1, 0)$	$\varphi(5) = (0, 1)$
6	2	9	$(0, 1)$	$\varphi(9) = (4, 1)$
7	1	5	$(0, 0)$	$\varphi(5) = (0, 1)$; back to identity
8	0	5	$(0, 1)$	$\varphi(5) = (0, 1)$; final state

Partial fold: $F_{+1} = (0, 1)$. Inverse: $(0, 1)^{-1} = (0, 1)$ (a reflection is its own inverse). TAJ-FC: $\varphi^{-1}(0, 1) = 0 + 5 = 5$.

The receiver, given TAJ $= 012345670$ and TAJ-FC $= 5$, places the digit 5 at position 0 and folds the ten-digit sequence; the result is $(0, 0)$, confirming validity.

Now suppose positions 4 and 5 (right-to-left) are transposed, changing the printed TAJ from 012345670 to 012354670. Recomputing the partial fold with the altered input yields $F_{+1} = (0, 0)$ and therefore TAJ-FC $= 0 \neq 5$, triggering the hard stop.

Online Resource 2: Reference Pseudocode

The two listings below translate the algebraic definitions into executable form. The group operations reduce to modulo-5 arithmetic plus a single XOR; the position-dependent permutation π_i uses two constant vectors ρ, ρ^{-1} and modulo-10 arithmetic.

Listing 1: Primitive operations on $S = \mathbb{Z}_5 \times \mathbb{Z}_2$ and the cycle-conjugate permutation π_i.

```
constant rho     := [0, 1, 4, 3, 8, 9, 6, 7, 2, 5]
constant rho_inv := [0, 1, 8, 3, 2, 9, 6, 7, 4, 5]

function Mul((a,b), (c,d)):
    sign := +1 if b = 0, else -1
    return ( (a + sign * c) mod 5,  b XOR d )

function Inv((a,b)):
    if b = 0: return ( (-a) mod 5, 0 )
    else:     return ( a, 1 )

function Phi(d):                         -- digit -> state
    if d < 5: return (d, 0)
    else:     return (d - 5, 1)

function PhiInv((a,b)):                  -- state -> digit
    if b = 0: return a
    else:     return a + 5
```

```
function Pi(d, i):                         -- position permutation
    x := rho_inv[d]
    y := (x + i) mod 10
    return rho[y]
```

Listing 2: TAJ-FC computation and vTAJ validation.

```
function ComputeTAJFC(t[0..8]):
    -- t[0] = rightmost digit
    S := (0, 0)                            -- identity
    for i from 0 to 8:
        S := Mul(S, Phi(Pi(t[i], i + 1)))
    return PhiInv(Inv(S))

function ValidateVTAJ(t[0..8], k):
    -- k = received TAJ-FC
    S := (0, 0)
    S := Mul(S, Phi(Pi(k, 0)))           -- position 0: TAJ-FC
    for i from 0 to 8:
        S := Mul(S, Phi(Pi(t[i], i + 1)))
    return S = (0, 0)
```

Online Resource 3: Additional Monte Carlo Data

Table S1 reports detection rates from the $n = 8{,}000$ pilot run. Results are consistent with the $n = 150{,}000$ run reported in the main text (Table I).

TABLE I
*

Table S1: Detection rates from Monte Carlo runs with $n = 8{,}000$ trials per configuration (seed 1). Wilson 95% intervals in the last two columns.

Error model	Codeword family	n	Det.	Rate	CI_{low}	CI_{high}
Single-digit subst.	legacy mod-10 (9)	8000	8000	1.0000	0.9995	1.0000
Single-digit subst.	fold explicit (9)	8000	8000	1.0000	0.9995	1.0000
Adjacent transpos.	legacy mod-10 (9)	8000	7059	0.8824	0.8751	0.8893
Adjacent transpos.	fold explicit (9)	8000	8000	1.0000	0.9995	1.0000
Random noise	legacy mod-10 (9)	8000	7543	0.9429	0.9376	0.9478
Random noise	fold explicit (9)	8000	8000	1.0000	0.9995	1.0000
Single-digit subst.	full vTAJ (10)	8000	8000	1.0000	0.9995	1.0000
Adjacent transpos.	full vTAJ (10)	8000	8000	1.0000	0.9995	1.0000
Random noise	full vTAJ (10)	8000	8000	1.0000	0.9995	1.0000

Figure S1 shows the convergence of Monte Carlo estimates between the two sample sizes. Point estimates are stable; Wilson intervals narrow by a factor of $\sqrt{150{,}000/8{,}000} \approx 4.3$, as expected from standard sampling theory.

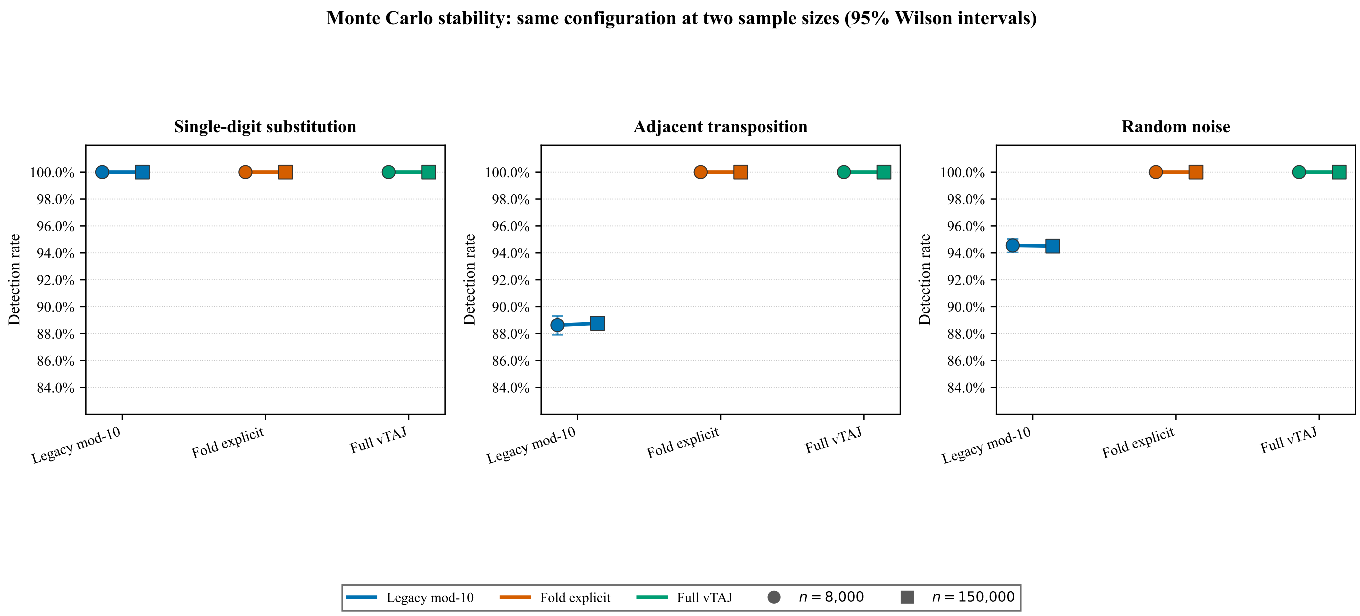

Fig. 1. *
Figure S1: Convergence of Monte Carlo estimates between $n = 8{,}000$ (circles) and $n = 150{,}000$ (squares). Thin horizontal ticks show Wilson 95% interval endpoints.