



**Digital
Forensics
Investigation
Research Laboratory**



Use of Generalized Hough Transform on interpretation of memory dumps

Paulo R. Nunes de Souza

Pavel Gladyshev

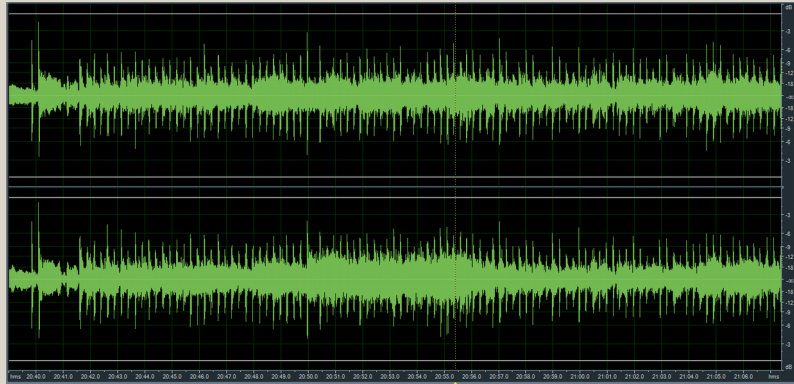


What is partial data?



- Incomplete data
- Partially corrupted data
- Data without required code/algorithm
- Data of unknown source

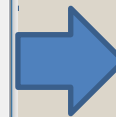
What does it look like?



C:\Users\Paulo Nunes\Dropbox\Paulo\PHD 2013\Conferences\Interpol DSC\presentation\unknown.dat - Notepad++

Address	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f	Dump
00002500	9d	2d	5d	51	89	cd	d2	46	f5	48	76	36	4a	81	80	39	.-]Q%IÒFðHv6J.€9
00002510	c7	27	5d	3a	fa	69	6b	a9	a7	9a	e0	f4	12	44	1a	a2	Ç'] :úik@S\$àø.D.ç
00002520	32	59	a0	43	b1	40	19	c8	1e	07	b7	8c	72	34	2f	67	2Y.C±@.È..ÇEr4/g
00002530	0a	e9	c0	91	f1	8e	1d	2b	f4	96	7e	b9	17	94	ba	b4	.éÀ`ñŽ.+ø-~¹."o´
00002540	a5	5a	ad	d0	4e	41	89	99	73	c8	27	dc	1f	9d	37	fd	¥Z-ÐNA%™sÈ'Û..7ý
00002550	41	e9	5a	0e	96	8e	d3	61	b7	09	56	4a	50	b1	4c	52	AeZ.-Žóa..VJP±LR
00002560	46	84	32	02	07	9c	f2	78	f3	a7	eb	7b	9d	26	6e	ad	F„2...øðxóSè{. &n-
00002570	b5	ad	af	e8	85	05	25	a4	bd	52	bf	6e	4a	9d	ed	dc	µ-`è...%ø¿R¿nJ.1ü
00002580	95	c3	30	cf	39	f6	d5	45	eb	2e	98	bd	4b	6d	82	ba	•Ä0İ9ðÖEè..½Km,°
00002590	c7	2d	47	72	8a	11	2d	4f	6f	2a	b1	b7	07	92	38	1c	Ç-GrŠ.-oo*±..´8.
000025a0	e3	fc	6a	65	d3	98	7d	d0	1f	59	ee	3d	2f	f6	7f	f8	šújeÓ.}Ð.Yi=/ð.ø
000025b0	8e	a2	77	fb	7a	33	05	3a	ee	67	86	76	c6	14	15	f8	Žčwûz3.:igtvE..ø
000025c0	e3	3b	bf	f6	d5	a7	e9	cf	ac	76	fa	9b	3c	82	be	39	ä;¿ðÖSéİ-vú><.,¾9
000025d0	27	ab	b8	6e	ed	18	93	14	e9	bf	80	ac	7f	b4	73	ae	'«.ní..".é¿@-..´s@
000025e0	85	95	63	f8	84	a1	8a	d3	0c	17	6b	75	55	35	15	60	...cø„jŠÓ..kuU5.`
000025f0	a5	32	48	82	4d	c6	7a	3c	8e	7c	f9	3c	df	be	a0	7b	¥2H,MEv<Ž¿ù>ß¾.{
00002600	a5	74	d7	8e	86	b6	45	72	32	8a	e9	41	dc	e0	6d	8d	¥t×Ž+¶Er2ŠéAŮàm.
00002610	41	1c	01	f2	72	32	7f	63	a1	e4	fc	30	7f	f7	31	be	A..òr2.c;äü0.+1¾
00002620	4e	8b	bc	45	d3	f1	55	8a	55	a8	96	14	5a	8a	76	11	N<¼EÓNUŠU"-..ZŠv.
00002630	07	48	8b	63	20	91	c7	39	ce	7f	03	52	97	43	5b	ab	.HcC`C9f...R-Cf«

Hex Edit View nb char: 314540 Ln: 0 Col: 9 Sel: 8 Hex BigEndian INS



Word puzzle



SCIENCE OF SECRETS

COELCJYVFODKGULGKPTD
MKVUCORSATELLITEOBHN
RZIJUWUONWWUFYLGGIW
TMDXMVIOLENTCRIMELTF
ELECTRONICENOHPTRAMS
TANRETNIVSJEMWGECEMAX
QFCJESPTEEGGMUMYMTAZ
UOEHDZNALPSAULEYCPKS
ZRCDALBLOODTINVVQISTY
QESCKOYETUAELMMTRRH
TNNBCFOGSLURCGRUOEEEN
ZSUOICOREFDCAGALIIDF
FIWSHBANVDEETQLTBWDX
XCHACPSQRVNSENVAILOA
KWP SHGOAADNMCGCITOPH
TBBPSPHXHIPATTERNSNH
WGSONOGRAMUIEWOLACUW
LCTERRORISMRMECOLIGI
TPUJHQBSLNXIHDECODE
IBJOFPREDATORVFQLUCG

ANTIBIOTIC	FLUORESCENT	SATELLITE
BACTERIA	FORENSIC	SAXOPHONE
BLOOD	GELATINOUS	SECRET AGENT
BRAINSTORM	GUNPOWDER	SMART PHONE
CHICKADEE	HARD DRIVE	SOLDIERS
COLLEGE	HARVEST	SONOGRAM
COMMUNICATE	INTERNET	STROLL
DECODE	INVESTIGATION	TERRORISM
E COLI	JELLYFISH	VIOLENT CRIME
ELECTRONIC	MERLIN	WIGGLY
EVIDENCE	PATTERNS	
FEROCIOUS	PREDATOR	

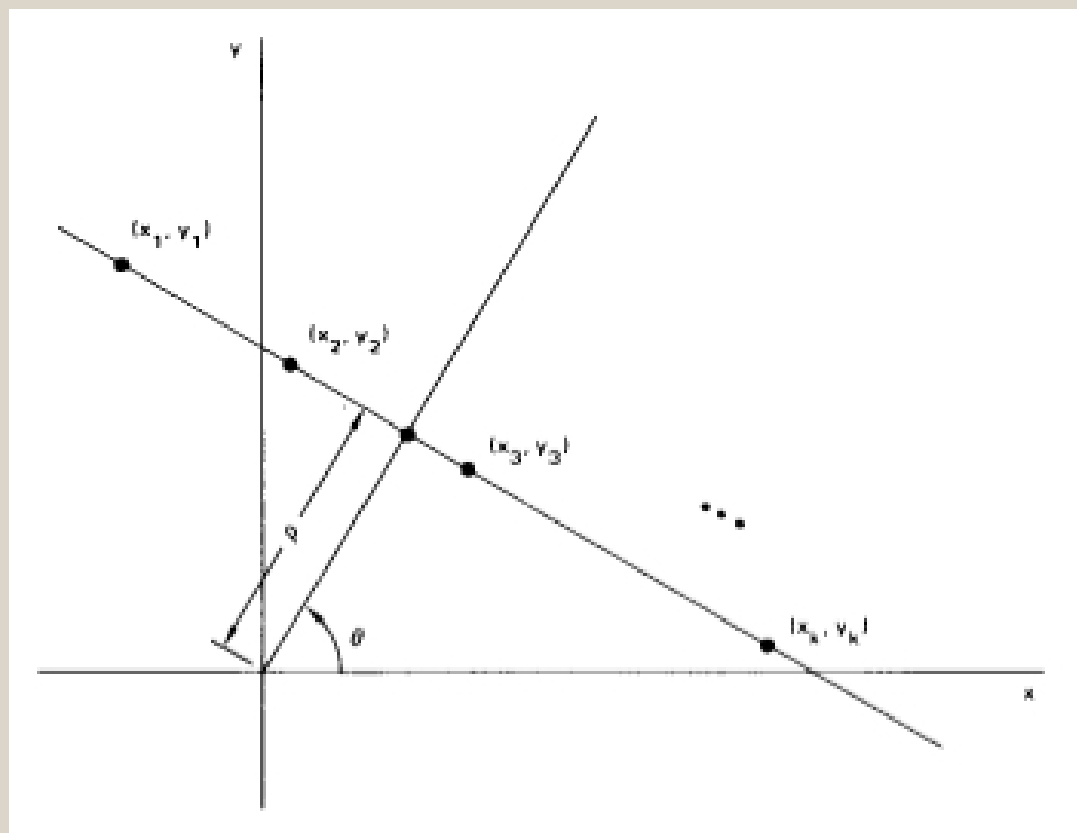
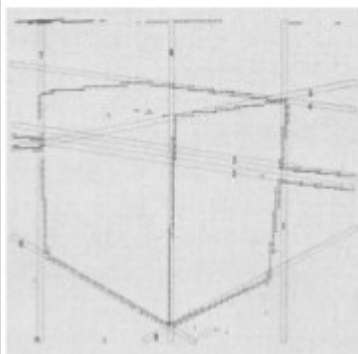
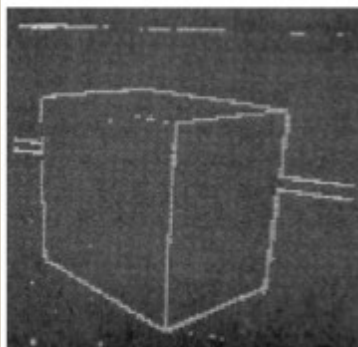
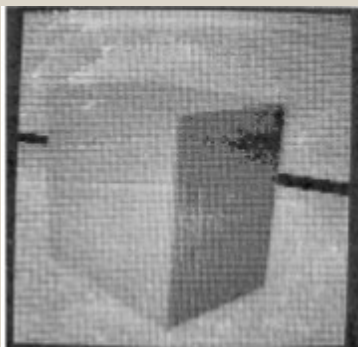


Requirements

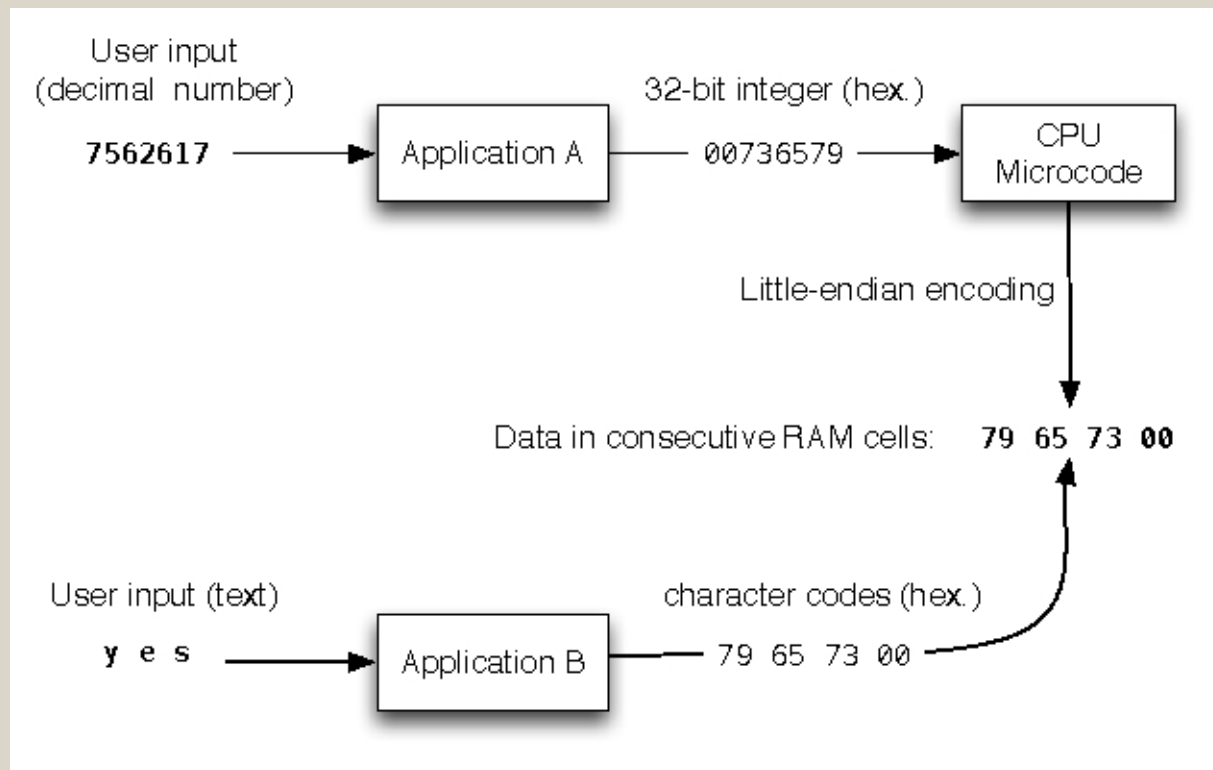


- Tolerance to noisy data;
- Tolerance to partial data corruption;
- Flexibility to define the structure being searched.
- Suitable candidate: **Hough Transform**

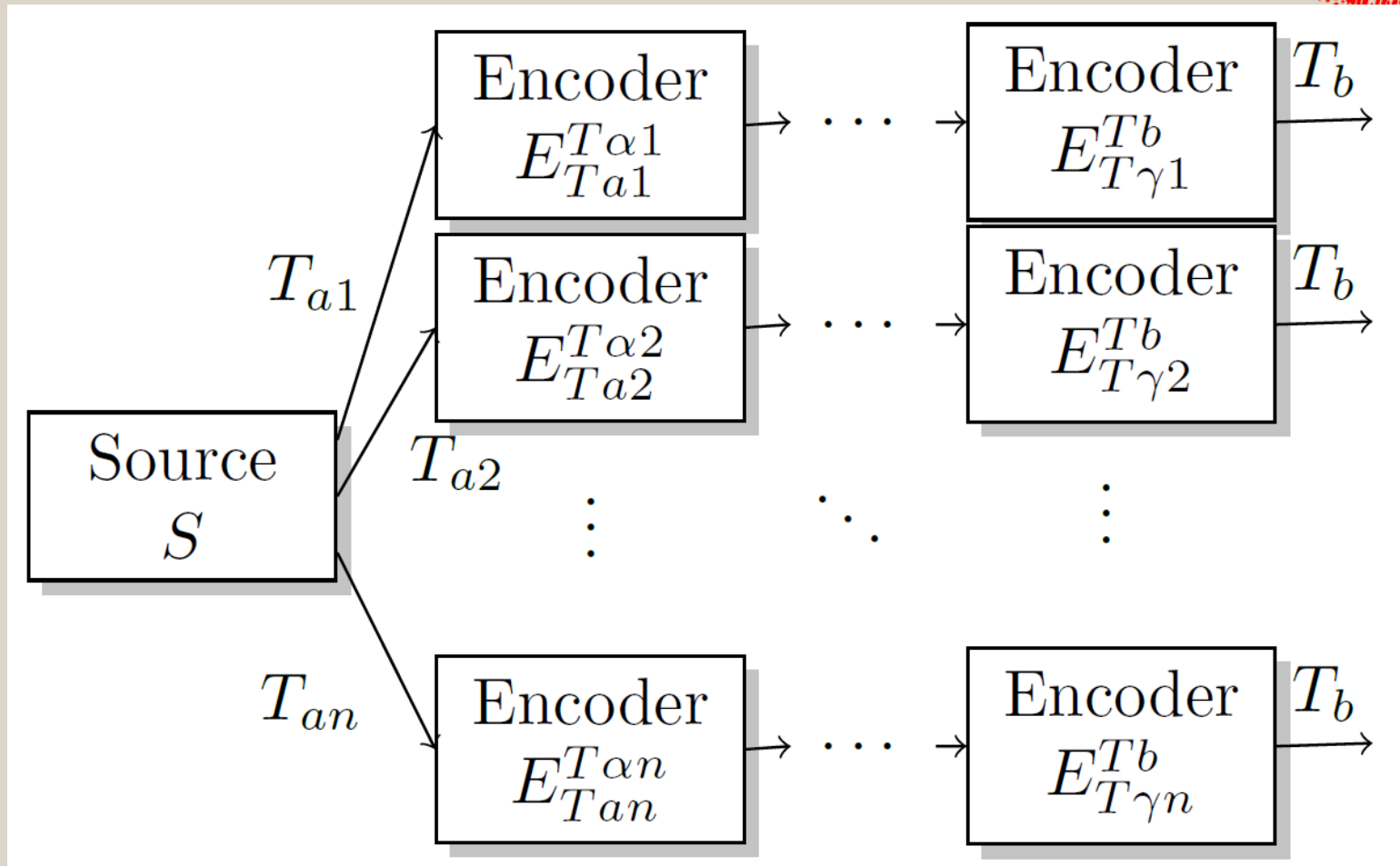
Hough Transform



Encoding

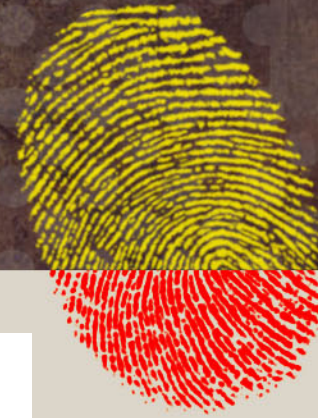


Model



$$T = \{T_{bin}, T_{int8}, T_{ascii}, \dots\}$$

Model



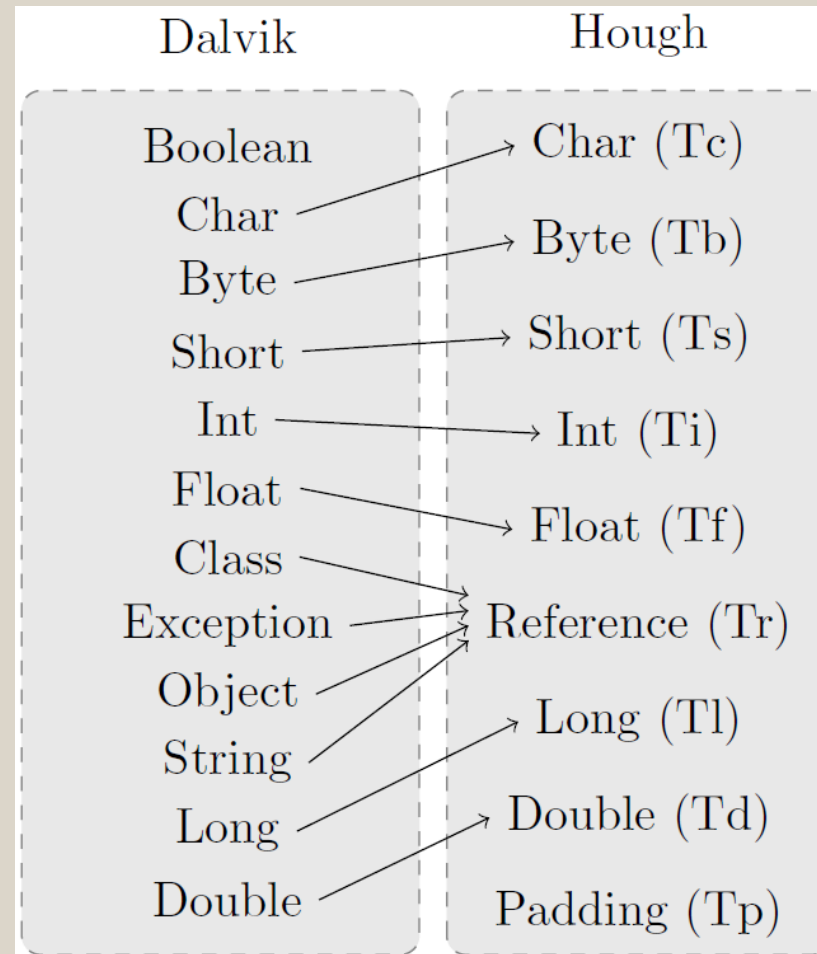
$$P_{E_{T_a}^{T_b}}(o|i) = \sum_{\substack{\sigma \in T_\alpha \\ \eta \in T_\beta \\ \mu \in T_\gamma}} P_{E_{T_a}^{T_\alpha}}(\sigma|i) \cdot P_{E_{T_\alpha}^{T_\beta}}(\eta|\sigma) \dots P_{E_{T_\gamma}^{T_b}}(o|\mu)$$

$$P_{E^{T_b}}(o|i, t) = \begin{cases} P_{E_t^{T_b}}(o|i) & \text{if } i \in t \\ 0 & \text{if } i \notin t \end{cases}$$

$$P_{SE^{T_b}}(w|k) = \prod_{j=0}^q \sum_{\forall \lambda \in k(j)} P_{E^{T_b}}(w(j)|\lambda, k(j))$$

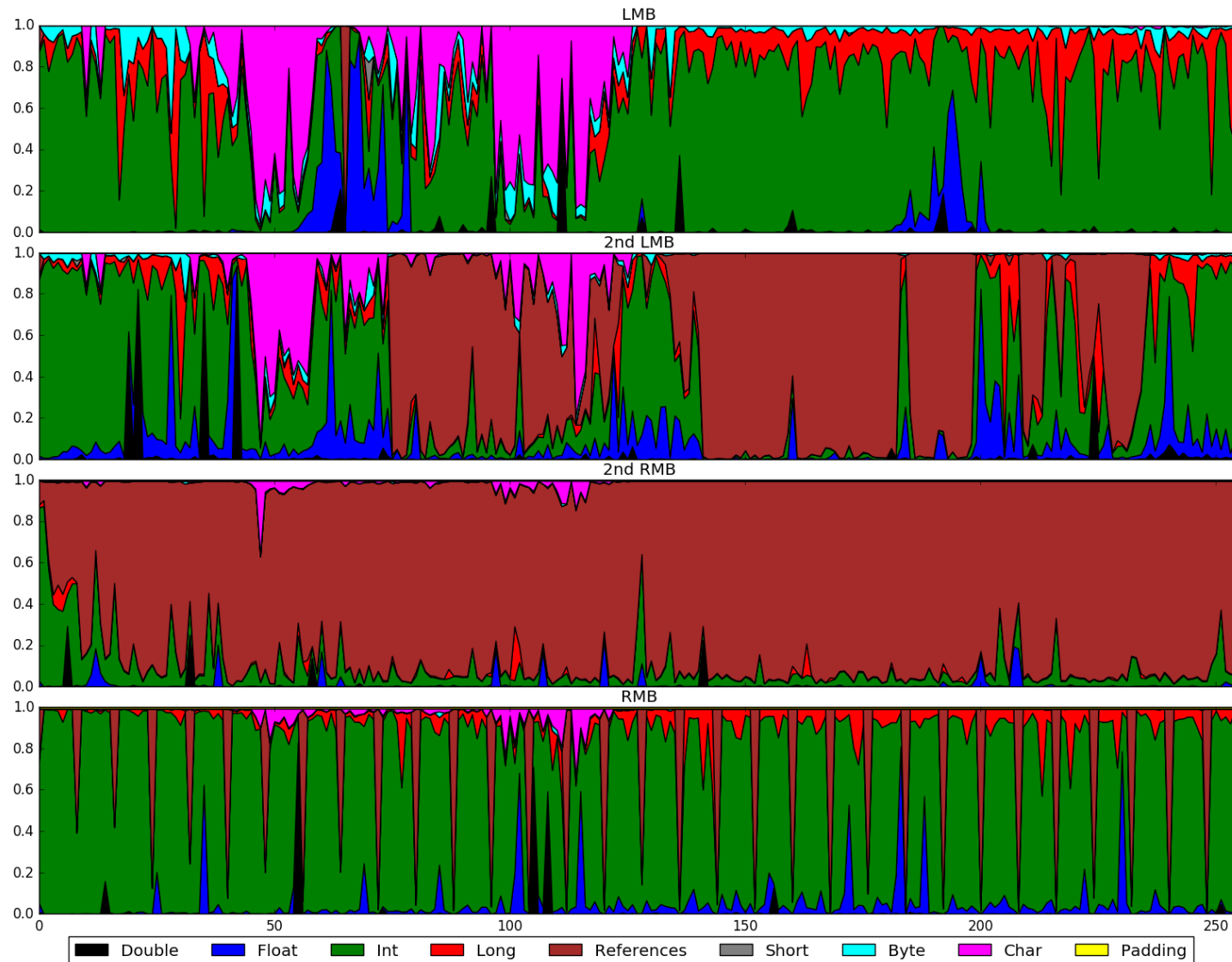
$$P_{SE}(k|w) = \frac{P_{SE}(w|k) \cdot P_{SE}(k)}{P_{SE}(w)}$$

Data types



$$T = \{Tbin, Tb, Tc, Td, Tf, Ti, Tl, Tp, Tr, Ts\}$$

Probabilities $P_{se}(k|w,b)$



R-table



Previous				Current				Next				
				k_0	k_1	k_2	k_3	k_4	k_5	k_6	k_7	...

(a) First word ($a = 0$)

Previous				Current				Next				
k_0	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_a	k_b	...

(b) Second word ($a = -4$)

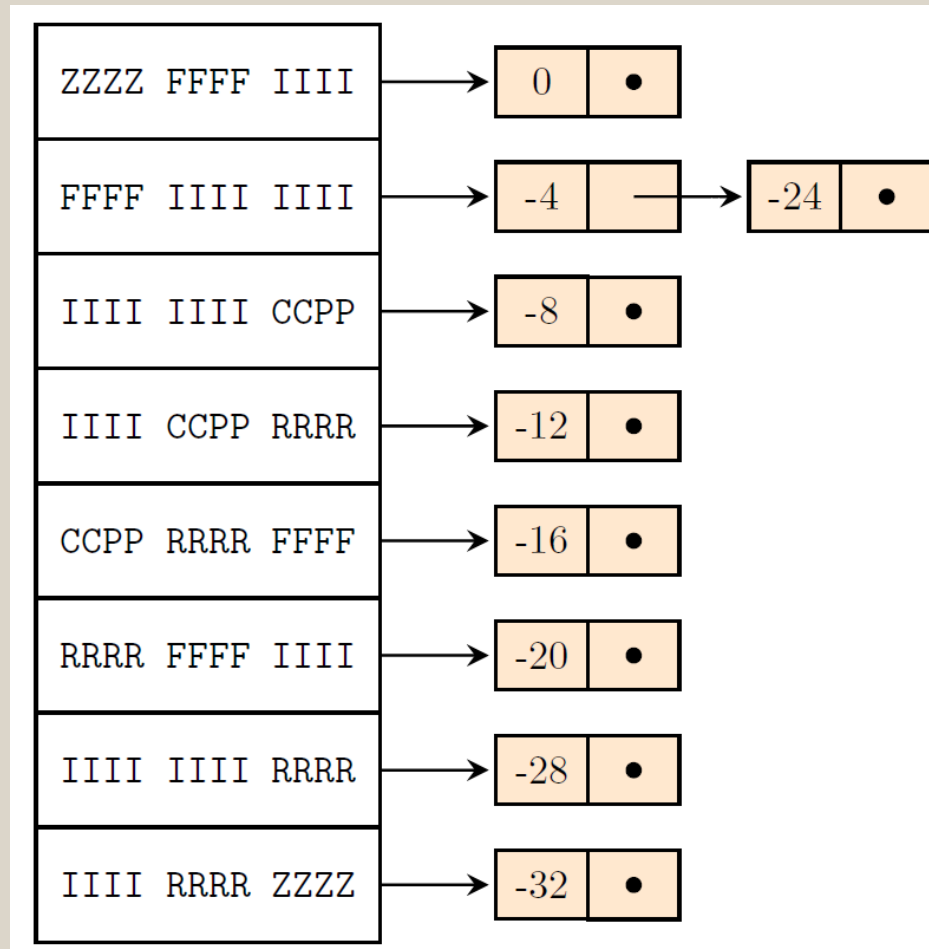
Previous				Current				Next				
...	k_8	k_9	k_a	k_b	k_c	k_d	k_e	k_f				

(c) Fourth and last word ($a = -12$)

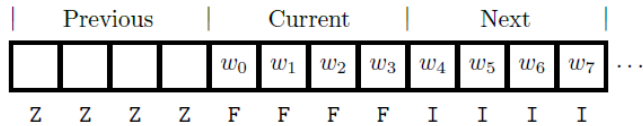
R-table



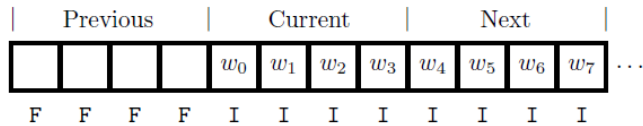
FFFF IIII IIII CCPP RRRR FFFF IIII IIII RRRR



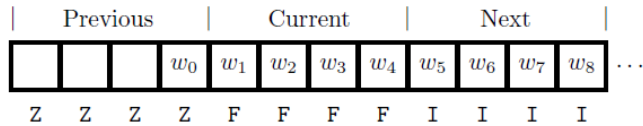
Accumulation



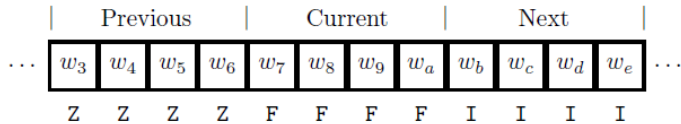
(a) Parsing process positioned on the first byte of the dump against the first R-table entry (ZZZZ FFFF IIII)



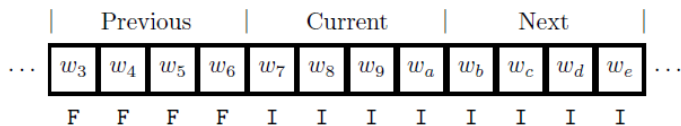
(b) Parsing process positioned on the first byte of the dump against the second R-table entry (FFFF IIII IIII)



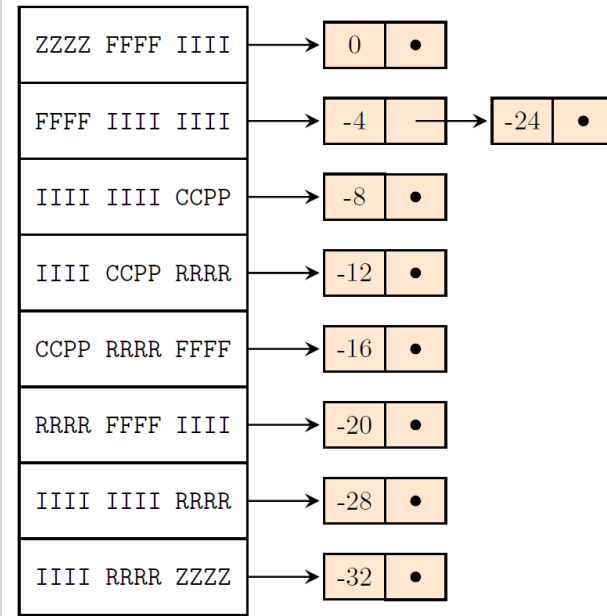
(c) Parsing process positioned on the second byte of the dump against the first R-table entry (ZZZZ FFFF IIII)



(d) Parsing process positioned on the eighth byte of the dump against the R-table first entry (ZZZZ FFFF IIII)



(e) Parsing process positioned on the eighth byte of the dump against the second R-table entry (FFFF IIII IIII)



$$\begin{aligned}
 C = & P_{SE}(F|w(3), 0) \cdot P_{SE}(F|w(4), 1) \cdot P_{SE}(F|w(5), 2) \cdot \\
 & P_{SE}(F|w(6), 3) \cdot P_{SE}(I|w(7), 0) \cdot P_{SE}(I|w(8), 1) \cdot \\
 & P_{SE}(I|w(9), 2) \cdot P_{SE}(I|w(a), 3) \cdot P_{SE}(I|w(b), 0) \cdot \\
 & P_{SE}(I|w(c), 1) \cdot P_{SE}(I|w(d), 2) \cdot P_{SE}(I|w(e), 3)
 \end{aligned}$$

If $C > C_t$, increment accumulation table at position $i+a$.

Peak detection



- Local maxima above threshold H
- Positions of dump where structure was identified

Tests



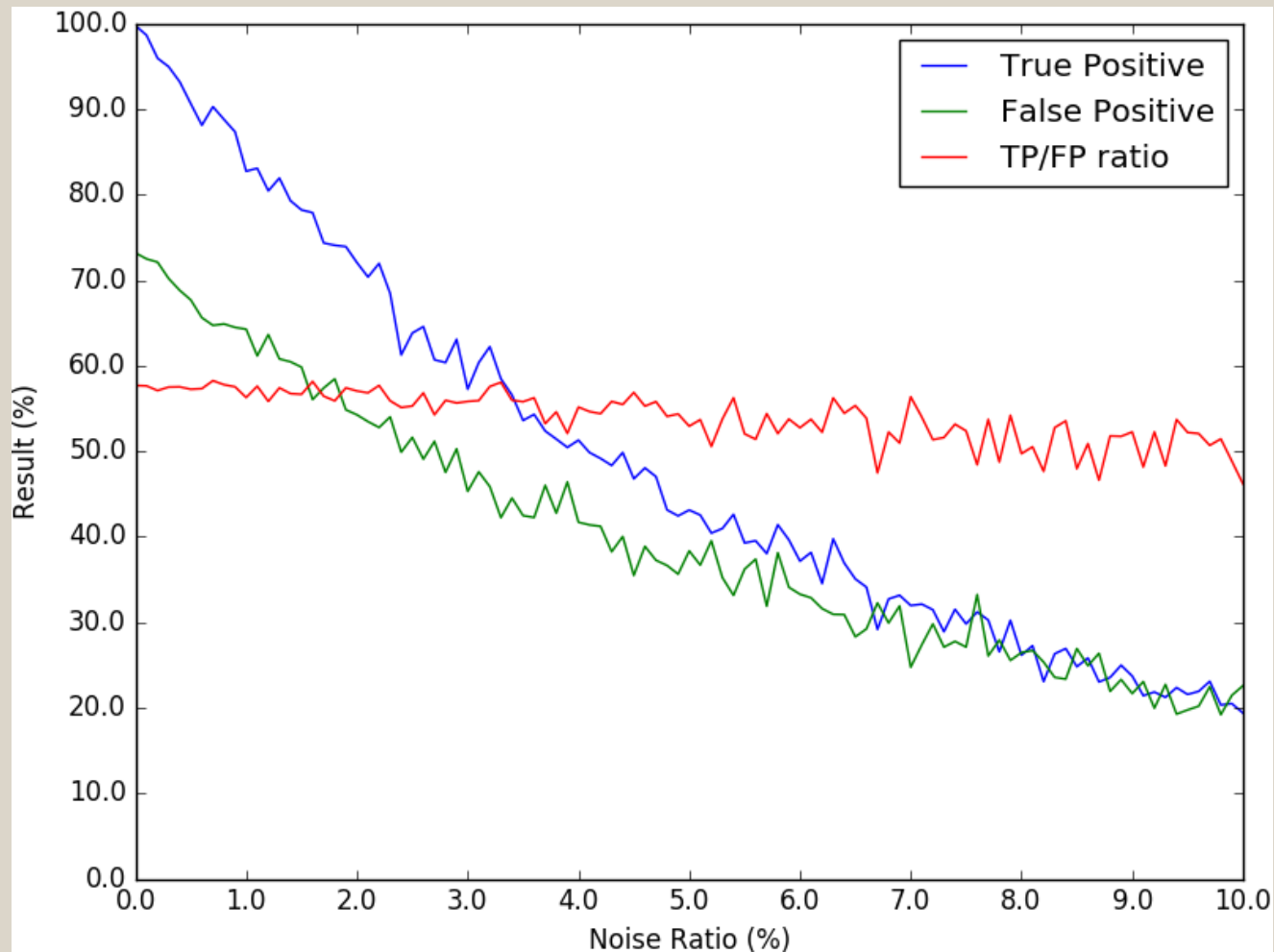
- Randomly created 100 distinct dumps
- Each dump with 4 kb in size
- Randomly choose one of the types in T
- Then a random value of the chosen type was inserted in the dump
- Repeat this until a 4kb dump was created
- The last step was to insert instances of the structure of interest across the previously created dumps.
- The position and quantity of those structures were also randomly chosen
- Each dump and the positions of the structure of interest were saved to respective files

Tests



- For each of the 100 dumps, another 100 versions of them were created
- One file for each level of added noise. Starting at 0% for each 0.1% step until reaching 10%
- At the end, we have 101 versions of each of the 100 dumps
- Total of 10100 test dumps

Results



Conclusions



- Tolerance to noisy data
- Flexibility to identify structure of interest
- The structure of interest was correctly spotted in 99.8% of the tests with no noise
- The structure of interest was correctly spotted in 20% of the cases with 10% of noise.
- The downside is the high false positive rate.
- Applicable beyond memory realm

Follow ups

- Hough-Forensic DSI
- PNG filecarving

The top of the slide features a dark, textured background with faint binary code (0s and 1s) and two overlapping fingerprints. The upper fingerprint is yellow and the lower one is red.

Thank you

