Masterkey attacks against free-text keystroke dynamics and security implications of demographic factors

Tim Van hamme, Giuseppe Garofalo, Davy Preuveneers, Wouter Joosen

DistriNet

A tale of high-tech wolves

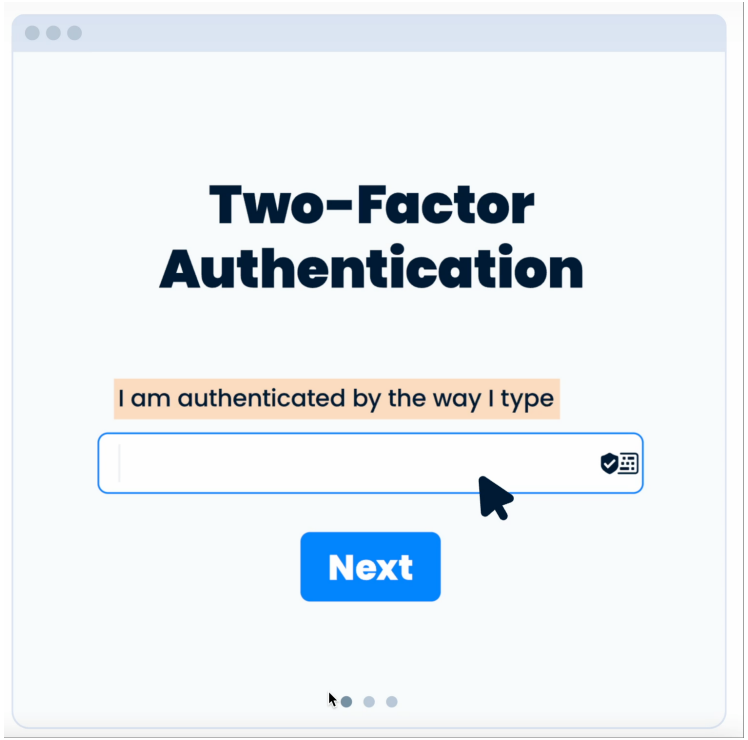
DistriNet

Masterkey

- Forge biometric data to improve wolves' **effectiveness**
- Behavioral authentication as perfect target
  - No secure hardware
  - No liveness detection

Keystroke biometrics

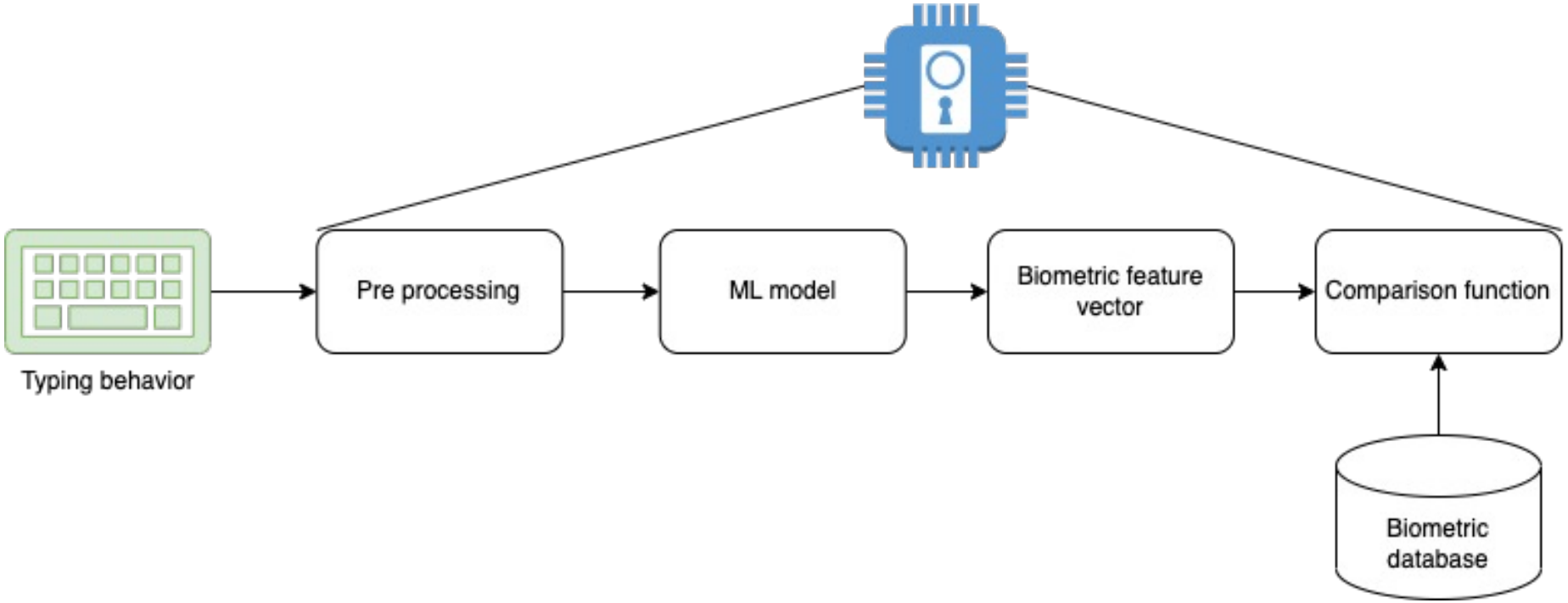
Use cases



- Fixed text
  - Second factor
- Free text
  - Second factor
  - Shopping cart check-out
  - Proctoring
  - Continuous authentication

Keystroke biometrics

System

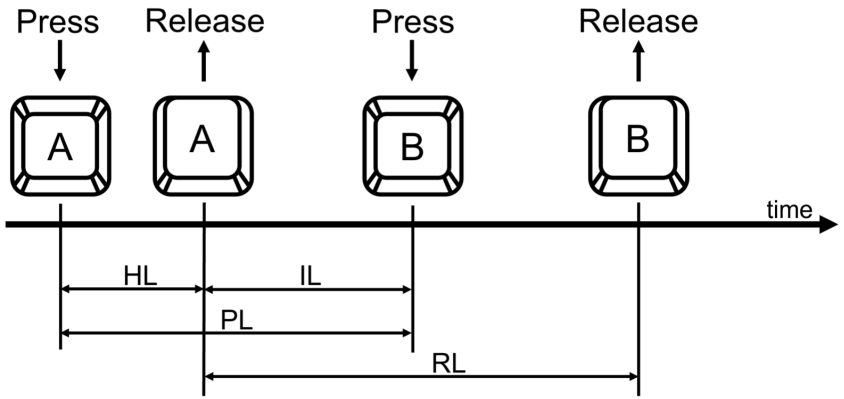


DistriNet

Keystroke biometrics

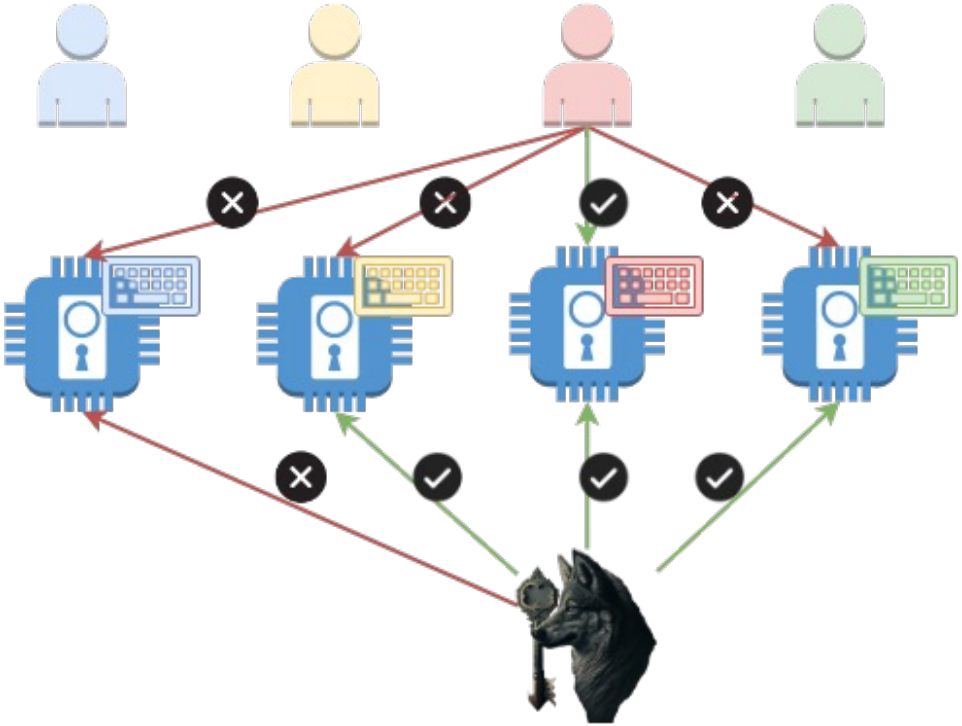
Feature extraction

- Hold latency
- Inter key latency
- Press latency
- Release latency



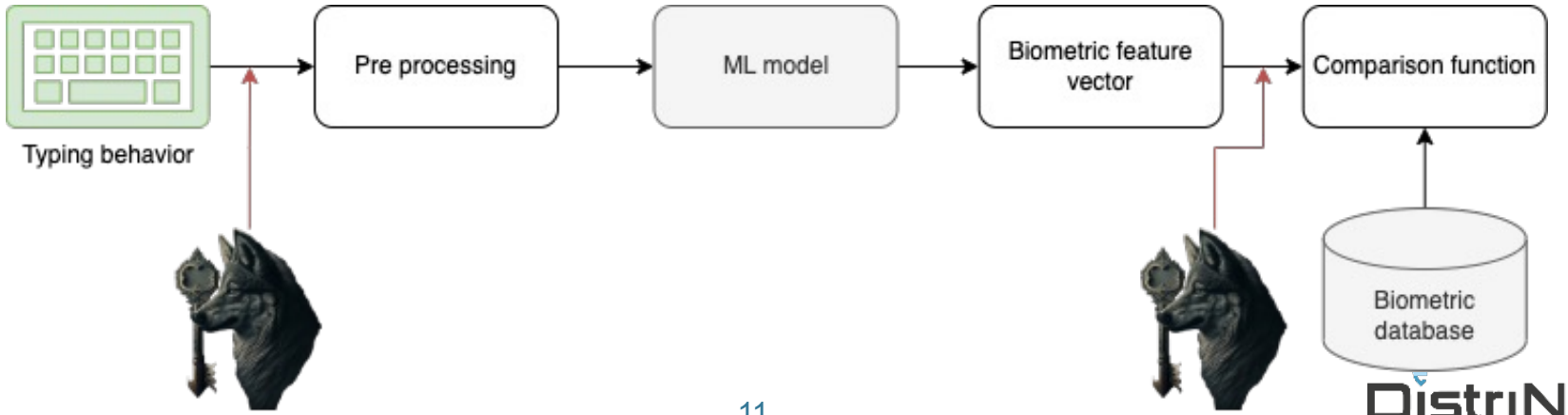
DistriNet

Masterkey attack against keystroke biometrics



Threat model

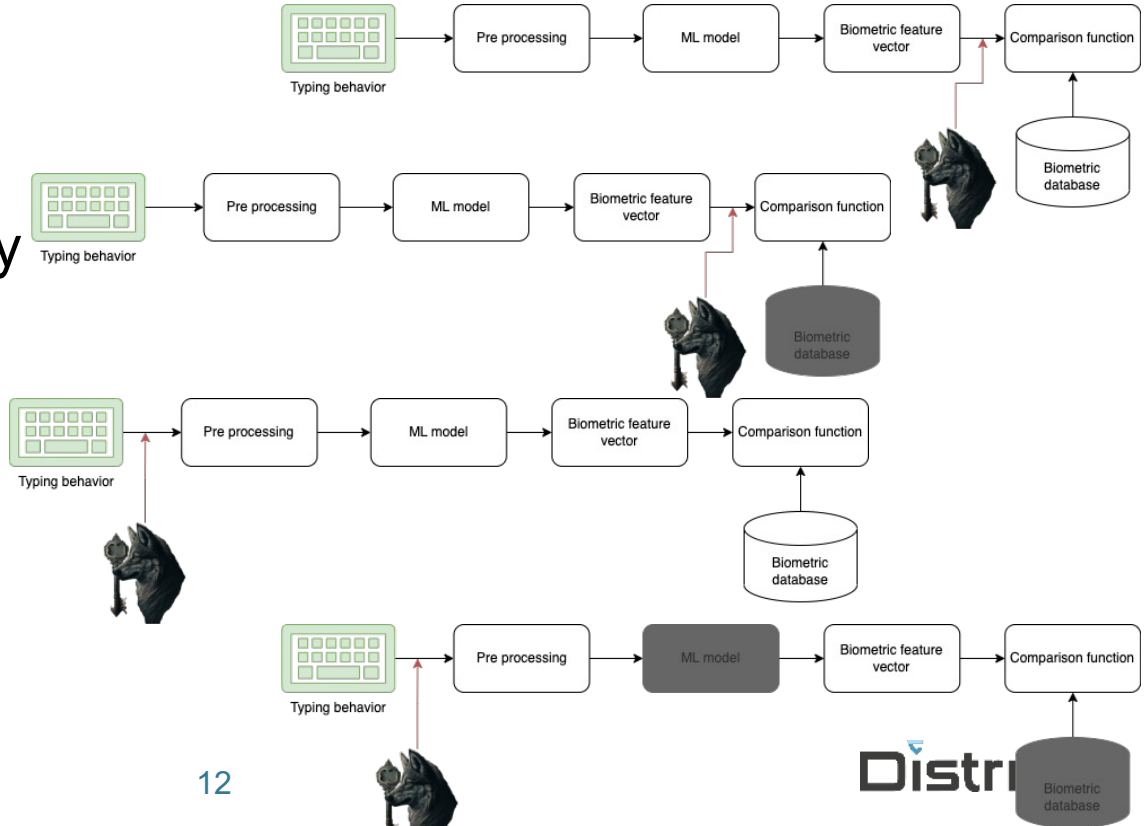
- Goal
  - Impersonate individual
- Capabilities
  - Inject biometric feature vectors
  - Inject keystroke timings
- Knowledge
  - Model (yes/no)
  - Enrolled population (yes/no)



DistriNet

Threat model

- Insider adversary
- Feature space adversary
- Input space adversary
- Pragmatic adversary



Attack methodology

- Feature space
  - Wolf samples
  - kMeans clustering
  - CMA-ES
- Key space
  - Wolf samples
  - CMA-ES

Brute force adversary

Background on guessing metrics

- Passwords and pincodes
  - Big corpus, extract most used passwords
  - Order p1> p2 > … > pn
- Min-entropy
  - Log2(p1)
- Beta success rate
  - Expected success rate (beta guesses)
  - Examples: Beta = 3 -> p1+p2+p3
- Alpha work factor
  - Number of guesses required to break at least a certain percentage of the account
  - Min(j | p1+p2+...+pj > alpha)
- Alpha guess work
  - Takes into account you can stop early
  - Weighted:
    1*p1 + 2*p2 + 3*(p3 + … + pn)

Brute force adversary

Background on guessing metrics

- Computer scientists like guessing difficulty as
  - Effective key length
  - Logarithmic scale
  - Bits of security
- The size N of the uniform distribution that yields same guessing value
  - p1 = p2 = … = pN
  - Log2(N)
  - 3 guesses 33% => [1/9]*9 => N = 9 (~3 bits of security)

Brute force adversary

Challenges for biometrics

- Challenges
  - Fuzzy matching
    - No exact match
    - Behavior similar enough
  - Uncountable events
    - Intepret as attack success rate

=> pi is the success rate of typing behavior i

  - Used to be occurences as a proxy of excepted success rate

=> What is the optimal guessing sequence?

  - NP-hard problem: maximum coverage problem

Brute force adversary

Biometric brute force adversary

	U1	…	Un	Sum
T1	1	…	1	12
T2	0	…	0	0
…		…		
Tm	1		0	4

Brute force adversary

Biometric brute force adversary

	U1	…	Un	Sum
T1	1	…	1	12
T2	0	…	0	0
…		…		
Tm	1		0	4

	U2	…	Un-1	Sum
T2	0	…	0	0
…		…		
Tm	0		1	4

12/n > 4/(n-12) > ...

Results

		H_∞	λ_3	λ_5	λ_{10}	$G_{0.25}$	$G_{0.5}$
Insider adversary	Wolf	4.38	4.69	4.8	5.02	4.91	5.28
	k-Means	3.93	4.21	4.28	4.46	4.27	4.46
	CMA-ES	3.66	3.82	3.92	4.18	3.86	4.06
Feature space adversary	Wolf	4.39	4.63	4.78	5.06	4.93	5.36
	k-Means	3.89	4.15	4.25	4.47	4.24	4.48
	CMA-ES	3.73	3.81	4.0	4.31	3.93	4.22
Input space adversary	CMA-ES	4.25	4.29	4.51	4.92	4.67	5.94
Pragmatic adversary	Wolf	4.91	5.42	5.47	5.6	5.65	5.82
	CMA-ES	6.64	7.1	7.74	8.67	8.91	8.91
Knowledge factors	Password [12]	6.5	/	/	9.1	17.6	21.6
	4-digit Pin [13]	4.75	5.22	5.5	5.91	6.32	8.78

Demographics analysis

		size	$\hat{\hat{H}}_{\infty}$	$\hat{\hat{\lambda}}_3$	$\hat{\hat{\lambda}}_5$	$\hat{\hat{\lambda}}_{10}$	$\hat{\hat{G}}_{0.25}$	$\hat{\hat{G}}_{0.5}$
Age	<18	33457	3.56	3.72*	3.84*	4.16*	3.77*	4.01*
	18-25	37790	3.39	3.69*	3.87*	4.23*	3.75*	4.1*
	25-35	24453	3.22	3.65*	3.86*	4.22*	3.71*	4.07*
	35-50	10410	3.11	3.5	3.7	4.07	3.49	3.82
	>50	3647	2.75	3.13	3.4	3.95	2.99	3.41
Sex	male	44864	3.46*	3.67*	3.87*	4.22*	3.75*	4.08*
	female	50788	3.46*	3.68*	3.86*	4.21*	3.74*	4.07*
Native language	en	92433	3.45	3.77	3.93	4.22	3.84	4.11
	es	2157	2.69	3.36	3.69	4.23	3.18	4.05
	hi	1402	2.59	3.0	3.31	4.02	2.8	3.24
	tl	2649	2.95	3.43	3.66	4.07	3.4	3.78
	zh	1721	3.35	3.67	3.84	4.18	3.69	4.0
Number of fingers	1-2	14050	2.35	3.02	3.43	4.01	2.71	3.44
	3-4	14149	2.82	3.11	3.41	3.96	2.98	3.42
	5-6	11770	3.21	3.53	3.73	4.13	3.55	3.91
	7-8	17487	3.73	3.89	4.04	4.33	3.97	4.27
	9-10	52473	2.92	3.4	3.64	4.11	3.36	3.8
Words per minute (speed)	<39.5	36767	2.3	2.74	3.06	3.82	2.52	2.81
	39.5-58.5	35931	2.98	3.21	3.47	3.98	3.16	3.54
	>58.5	37232	2.16	2.65	2.95	3.76	2.38	2.67

Conclusion

- Masterkey attack against keystroke biometrics
- New security metric
  - Comparison with passwords and pin codes
  - Demographic analysis

Questions?

https://distrinet.cs.kuleuven.be/