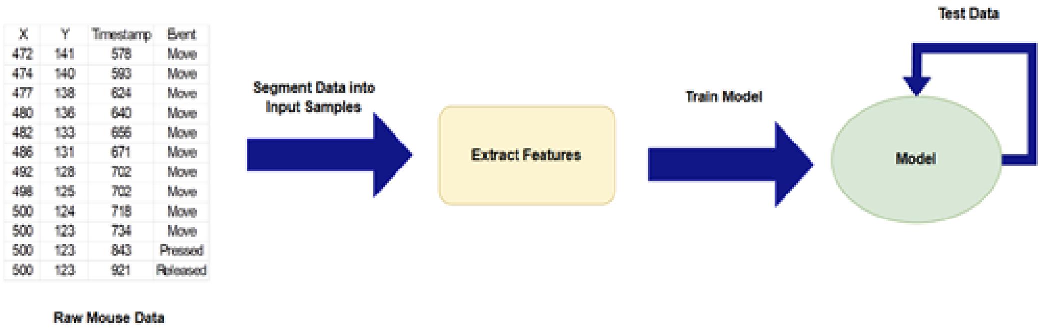
Behavior-Aligned Mouse Dynamics Segmentation

Charles Devlen, *Daqing Hou*

Mouse Dynamics for authentication

- As digital systems become more prevalent, so does the risk of personal data breach [1-3]
- Adoption of additional factors remains limited despite their effectiveness
- Mouse dynamics authenticates users passively, **without additional effort from user**
- **Core assumption:** Users have distinctive patterns of mouse use that can be leveraged for authentication

Typical Mouse Dynamics Workflow



X	Y	Timestamp	Event
472	141	578	Move
474	140	593	Move
477	138	624	Move
480	136	640	Move
482	133	656	Move
486	131	671	Move
492	128	702	Move
498	125	702	Move
500	124	718	Move
500	123	734	Move
500	123	843	Pressed
500	123	921	Released

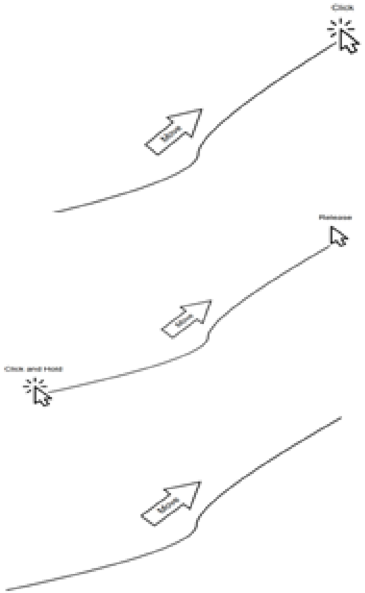
[1] FBI, “U.S. charges russian fsb officers and their criminal conspirators for hacking yahoo and millions of email accounts,” https://www.justice.gov/opa/pr/us-charges-russian-fsb- officers-and-their-criminal-conspirators-hacking-yahoo-and-millions, Mar 2017
[2] ——, “Chinese military personnel charged with computer fraud, economic espionage and wire fraud for hacking into credit reporting agency equifax,” https://www.justice.gov/opa/pr/chinese-military-personnel-charged-computer-fraud-economic-espionage-and-wire-fraud-hacking, Feb 2020
[3] J. Dhaliwal, “26 billion records released in “the mother of all breaches”,” https://www.mcafee.com/blogs/internet-security/26-billion-records-released-the-mother-of-all-breaches/, Jan 2024

Segmentation - How samples be separated from the raw data

- What constitutes a sample?
    - Diverse methods have been utilized in prior work for mouse data segmentation
    - Most prior work treats segmentation as a preprocessing heuristic instead of an explicit representation decision
- Segmentation is important as it determines
    - What behavioral information is preserved
    - The context of the sequence
    - What patterns can be learned

Definitions of Action

Action Type	Definition
Point-and-click	A series of mouse move events ending with a click
Drag-and-drop	A mouse button press event followed by a series of move events, ending with a mouse button release
Pure Movement	A series of move events not containing click events
Compound Action	A combination of pure movement and point-and-click or drag-and-drop action

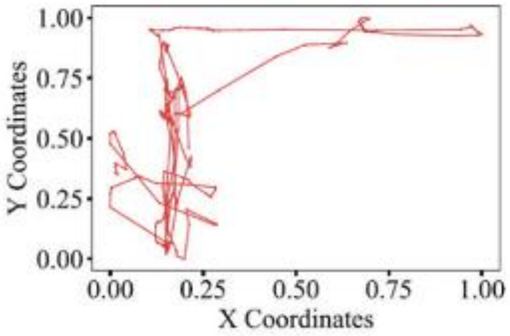


Pure Movement:

- Hovering over UI elements
- Tracing text while reading
- Looking around in a game

Existing Segmentation Approaches

Method	Idea	Example Paper
Action-Driven	Extract actions such as point-and-click, drag-and-drop, and compound actions by parsing events	Shen et al. [1] extract single click, double click, common movement (a.k.a. pure movement), point-and-click, drag-and-drop, and pause actions.
Time-Difference Thresholding	Split the raw data based on pauses between events in the data (e.g., 1 second)	Tan et al. [2] separated data into segments using a 500ms time-difference threshold
Windowing	Fixed windows of N events, either overlapping or non-overlapping	Pusara and Brodley [3] sliced mouse movement data into windows of 100, 200, 400, 600, 800, or 1000 events, determined by which size minimized the error rate per user.
Hybrid	Combination of the above	Antal et al. [4] extract point-and-click and drag-and-drop actions, then further extract mouse movement (a.k.a. pure movement) based on a time-difference threshold of 10 seconds.



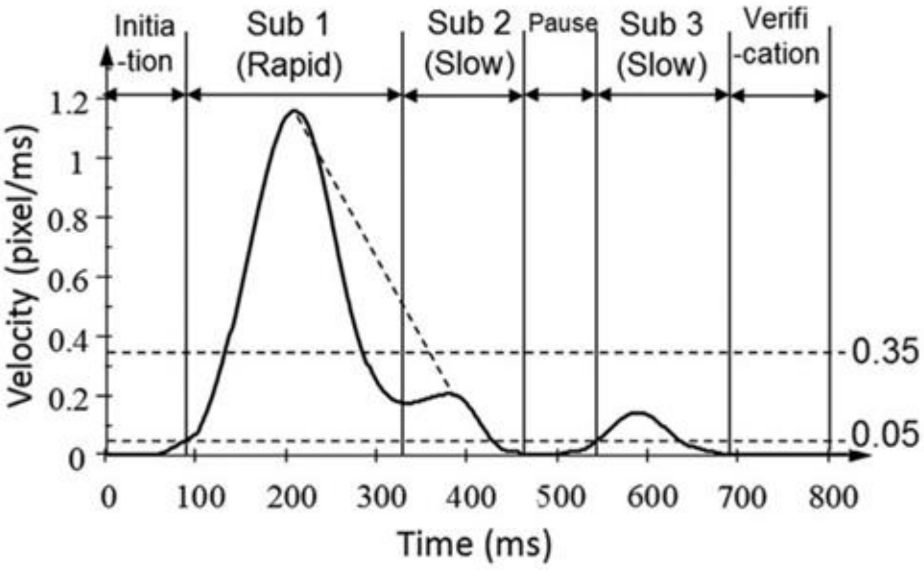
[1] C. Shen, Z. Cai, and X. Guan, “Continuous authentication for mouse dynamics: A pattern-growth approach,” in IEEE/IFIP Intl. Conference on DSN ’12. IEEE, 2012, pp. 1–12.
[2] Y. X. M. Tan, A. Binder, and A. Roy, “Insights from curve fitting models in mouse dynamics authentication systems,” in 2017 IEEE Conference on Application, Information and Network Security (AINS). IEEE, 2017, pp. 42–47.
[3] M. Pusara and C. E. Brodley, “User re-authentication via mouse movements,” in Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security, 2004, pp. 1–8.
[4] M. Antal and L. Denes-Fazakas, “User verification based on mouse dynamics: a comparison of public data sets,” in 2019 IEEE 13th International Symposium on Applied Computational Intelligence and Informatics (SACI), 2019, pp. 143–148.

Limitations of Existing Methods

- Most prior works focus on action-driven approaches
  - While actions such as point-and-click and drag-and-drop are extracted, actions consisting only of motion are not extracted
  - These “pure movement” actions are often merged with other action types
- **Pure movement** makes up a substantial portion of interaction behavior:
  - Hovering over UI elements
  - Tracing text while reading
  - Looking around in a game
- We found pure movement can represent **more than half of user activity**

Human Computer Interaction Principles

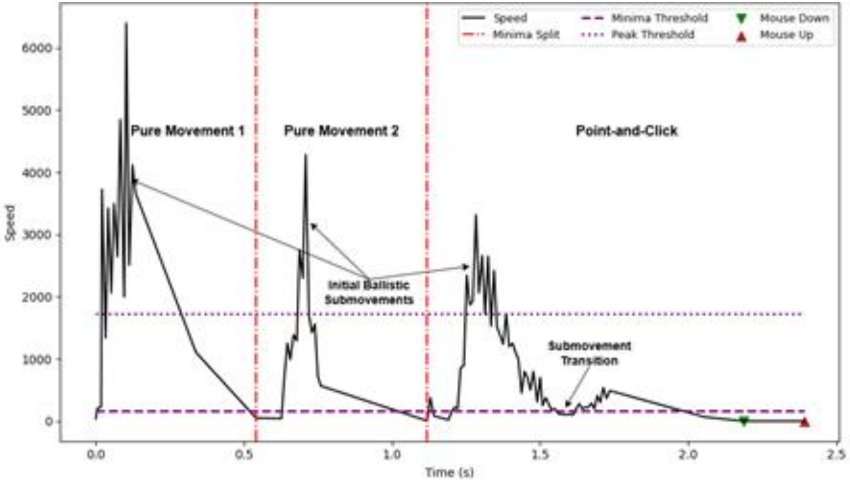
- Fitts' law is a foundational HCI principle that describes the structure of **hand movement target acquisition** tasks [1]
- Exploring how Fitts' law applied to mouse data, Chen et al. [2] revealed the structure of mouse target acquisitions
  - First submovement is **ballistic** in nature
  - Following submovements (if any) are **smaller and corrective** approaching the target
  - **A verification** period, either a click or sustained low velocity



[1] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement." Journal of experimental psychology, vol. 47, no. 6, p. 381, 1954.
[2] Y. Chen, E. R. Hoffmann, and R. S. Goonetilleke, "Structure of hand/mouse movements," IEEE Transactions on Human-Machine Systems, vol. 45, no. 6, pp. 790–798, 2015.

Proposed Segmentation Pipeline

1) Extract actions by parsing events
2) Calculate the speed at each point in the action
3) Define two thresholds
    a) Minima threshold: Signals very low speed
    b) Peak threshold: Signals ballistic submovements
4) Between each consecutive pair of ballistic peaks, find candidate velocity minimas
5) For each candidate minima, from left to right:
    a) Look ahead 500 ms
    b) If a ballistic peak does not appear, keep searching
    c) Otherwise, select the minima as the **split point**



Datasets and architecture

- We evaluated our behavior-aligned segmentation method across three publicly available datasets
  - **Balabit** [1] - Collected from 10 users as they completed their **everyday work on a remote server**
  - **ChaoShen** [2] - Collected from 28 users in a **laboratory setting, unconstrained computing tasks**
  - **DFL** [3] - Collected from 21 users on their **personal devices, unconstrained computing tasks**
- Together, these datasets represent a wide range of tasks and operating conditions
- Input Features - dx, dy, dt
  - From these features, the model can potentially learn more complex features such as velocity or change in angle
- Architecture - **Bidirectional LSTM**
  - Capture context based on past and future timestamps
  - Able to handle variable length sequences
  - Natural choice of representation for the time-series mouse data

[1] A. F¨ul¨op, L. Kovacs, T. Kurics, and E. Windhager-Pokol, “Balabit mouse dynamics challenge data set,” https://github.com/balabit/Mouse-Dynamics-Challenge, 2016, Last Accessed: 2026-03-08.
[2] C. Shen, Z. Cai, and X. Guan, “Continuous authentication for mouse dynamics: A pattern-growth approach,” in IEEE/IFIP Intl. Conference on DSN ’12. IEEE, 2012, pp. 1–12.
[3] M. Antal and L. Denes-Fazakas, “User verification based on mouse dynamics: a comparison of public data sets,” in 2019 IEEE 13th International Symposium on Applied Computational Intelligence and Informatics (SACI), 2019, pp. 143–148.

Current State-of-the-Art

- Mazumdar and Sundaram [1] as SOTA
  - utilized a **vision transformer** architecture and **recurrence plot images**
  - window-based segmentation strategy: 25% overlapping windows of events were used for training, and non-overlapping windows were used for testing
  - evaluated windows of 300, 450, and 600 events

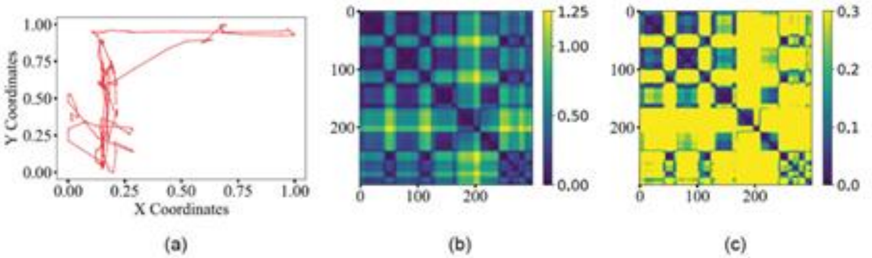


Fig. 1. In sub-figure (a), we show the mouse coordinate plot of 300 events from session 0735651357 of user 16 of the Balabit dataset. Sub-figures (b) and (c) correspond to the distance matrix and the recurrence plot. A threshold of $\epsilon = 0.30$ is applied on the distance matrix to obtain the recurrence plot representation. The size of distance matrix and recurrence plot is of 300×300.

[1] K. Mazumdar and S. Sundaram, "A mouse dynamics authentication system with a recurrence plot image representation and a vision transformer framework," IEEE Transactions on Information Forensics and Security, vol. 20, pp. 6895–6909, 2025.

Segmentation Methods to be Compared

- We examine the performance of three other segmentation methods against our own
  - **Action-Driven** - Point-and-click, drag-and-drop, and compound actions are parsed from the data
  - **Hybrid** - This approach segments the data into actions through event parsing, then further subdivides based on a global time difference threshold of 10 seconds [1]
  - **Window-Based** - We adopt the window-based approach used in the prior SOTA work [2]

- For the segmentation methods that derive actions, the same action parsing algorithm is used for consistency across experiments

[1] M. Antal and L. Denes-Fazakas, "User verification based on mouse dynamics: a comparison of public data sets," in 2019 IEEE 13th International Symposium on Applied Computational Intelligence and Informatics (SACI), 2019, pp. 143–148.
[2] K. Mazumdar and S. Sundaram, "A mouse dynamics authentication system with a recurrence plot image representation and a vision transformer framework," IEEE Transactions on Information Forensics and Security, vol. 20, pp. 6895–6909, 2025.

Dataset Composition

- Behavior-aligned segmentation resulted in significantly more actions than other methods
- While hybrid segmentation increased the number of samples over action driven, the global time-difference threshold was too rigid for capturing pure movement
- The less constrained datasets are composed of more pure movement, likely due to additional tasks such as exploratory navigation or gaming

Dataset	# Samples **Behavior-Aligned**	# Samples **Hybrid**	# Samples **Action-Driven**
Balabit	167,937 (39.98%)	106,128 (5.02%)	100,797
ChaoShen	1,267,472 (63.75%)	490,578 (6.34%)	459,485
DFL	2,685,164 (60.34%)	1,108,519 (3.94%)	1,064,809

Single-Sample Results

- Behavior-aligned segmentation performs worse than both hybrid and action-driven segmentation
- Why?
  - Behavior-aligned samples contain less events on average, so the observed window of behavior is smaller
- Tradeoff: Behavior-aligned segmentation has less discriminable single samples, but nearly half the events of action-driven segmentation
- Is the tradeoff worthwhile for a more rapid authentication decision?

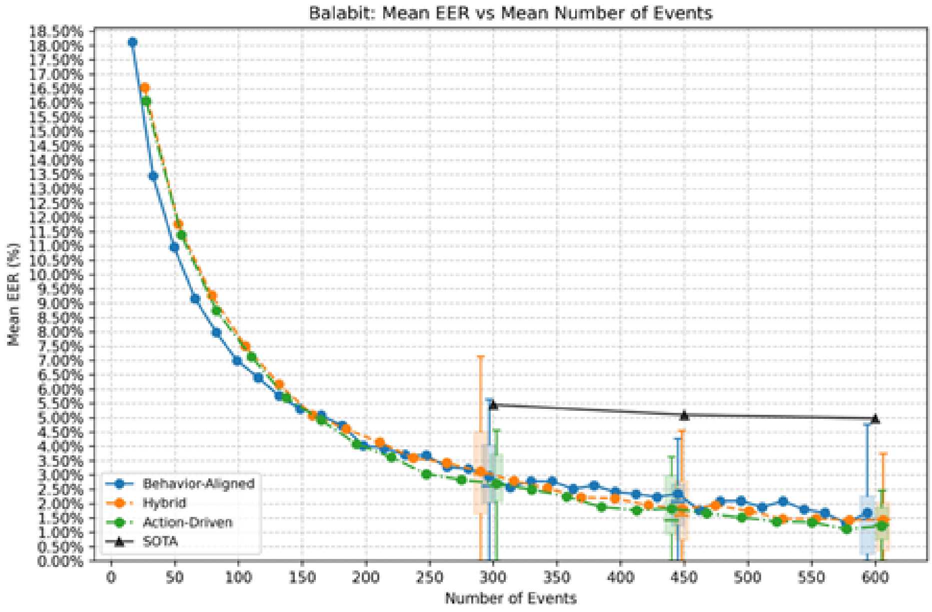
Dataset	Behavior-Aligned			Hybrid			Action-Driven		
	EER (%)	AUC	Mean # Events	EER (%)	AUC	Mean # Events	EER (%)	AUC	Mean # Events
Balabit	18.12	0.8978	16	16.53	0.9061	26	16.06	0.9123	28
ChaoShen	11.70	0.9435	36	10.69	0.9492	90	10.48	0.9496	95
DFL	6.55	0.9798	38	4.95	0.9873	71	4.66	0.9885	73

Exploring the Tradeoff - Score-Level Fusion

- To determine the relative performance of each method, comparisons should be made where the number of events is equal
- To this end, we adopt a simple score-level fusion strategy that averages the model output scores for N consecutive samples
- By multiplying the number of trajectories fused by the mean number of events in a single sequence, we can compare each method's EER as function of number of events
- By interpolating per-user EER over the observed range of events, we enable paired observations allowing for testing of significant difference

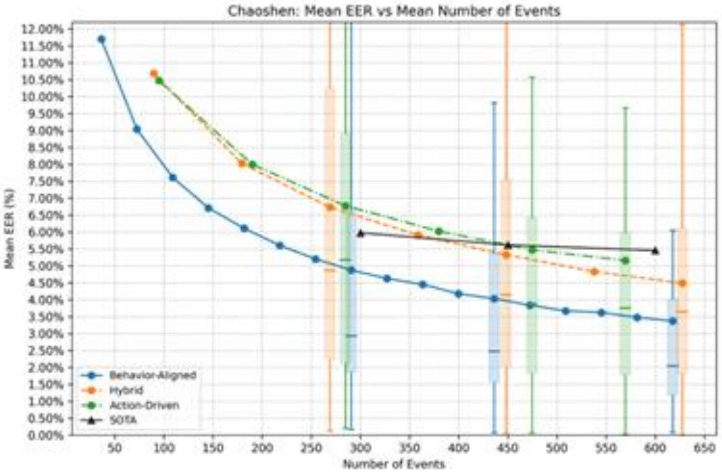
Score-Level Fusion Results: Balabit

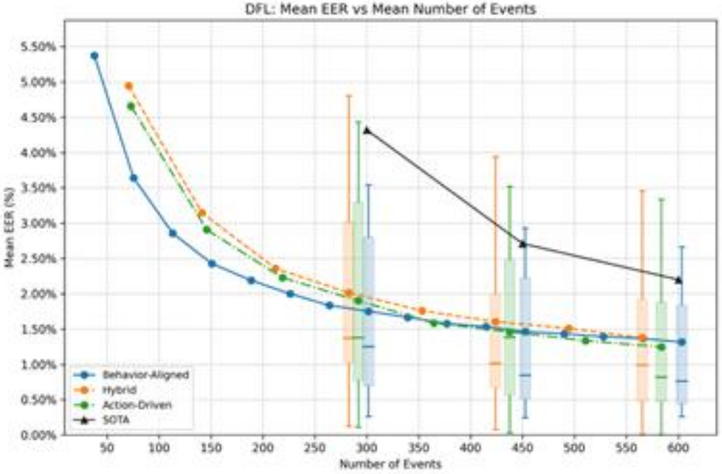
- Improvements in mean EER are apparent at earlier event counts, but taper off as more samples are aggregated
  - **Statistically significant differences** between action-driven and behavior-aligned segmentation were observed **up to 50 events**, with borderline significance at 30 events
- This relationship implies that behavior-aligned segmentation is more beneficial in the early decision regime
- Once sufficient behavioral evidence is accumulated, the advantage of behavior aligned segmentation diminishes



Score-Level Fusion Results: ChaoShen and DFL

- Behavior-aligned segmentation demonstrates a clear advantage over the other methods for ChaoShen and DFL
- Similarly to Balabit, differences between mean EER across methods decrease as more sequences are aggregated
- Why is the difference more pronounced than in the Balabit experiments?
  - Balabit has the least amount of pure movement extracted of the three datasets
- Statistical significance between action-driven and behavior-aligned segmentation was observed **up to 310 events** for ChaoShen and **160 events** for DFL





Score-Level Fusion Results - Takeaways

- A clear relationship was demonstrated between the amount of pure movement present and the performance benefit of behavior aligned segmentation
  - Balabit has the least pure movement and had the least benefit, whereas ChaoShen had the largest benefit and largest proportion of pure movement

- Behavior-aligned segmentation is most useful in the early decision regime
  - Impactful for authentication, where detecting misuse of a system as soon as possible is desirable

Conclusions

- Segmentation should be treated as a representation decision instead of preprocessing heuristic
- Pure movement matters, ~40-60% of observed behavior
- Behavior-aligned segmentation improves performance when behavioral evidence is limited
- State-of-the-art performance across three publicly available datasets
- Behavior-aligned segmentation generalizes across architectures

Future Directions

- User specific thresholds
- Learnable threshold parameters
- Explore time-series transformer architectures
- Explore long term drift and hardware dependence