Simultaneous Detection and Characterization of Time Delay Attack in Cyber-Physical Systems

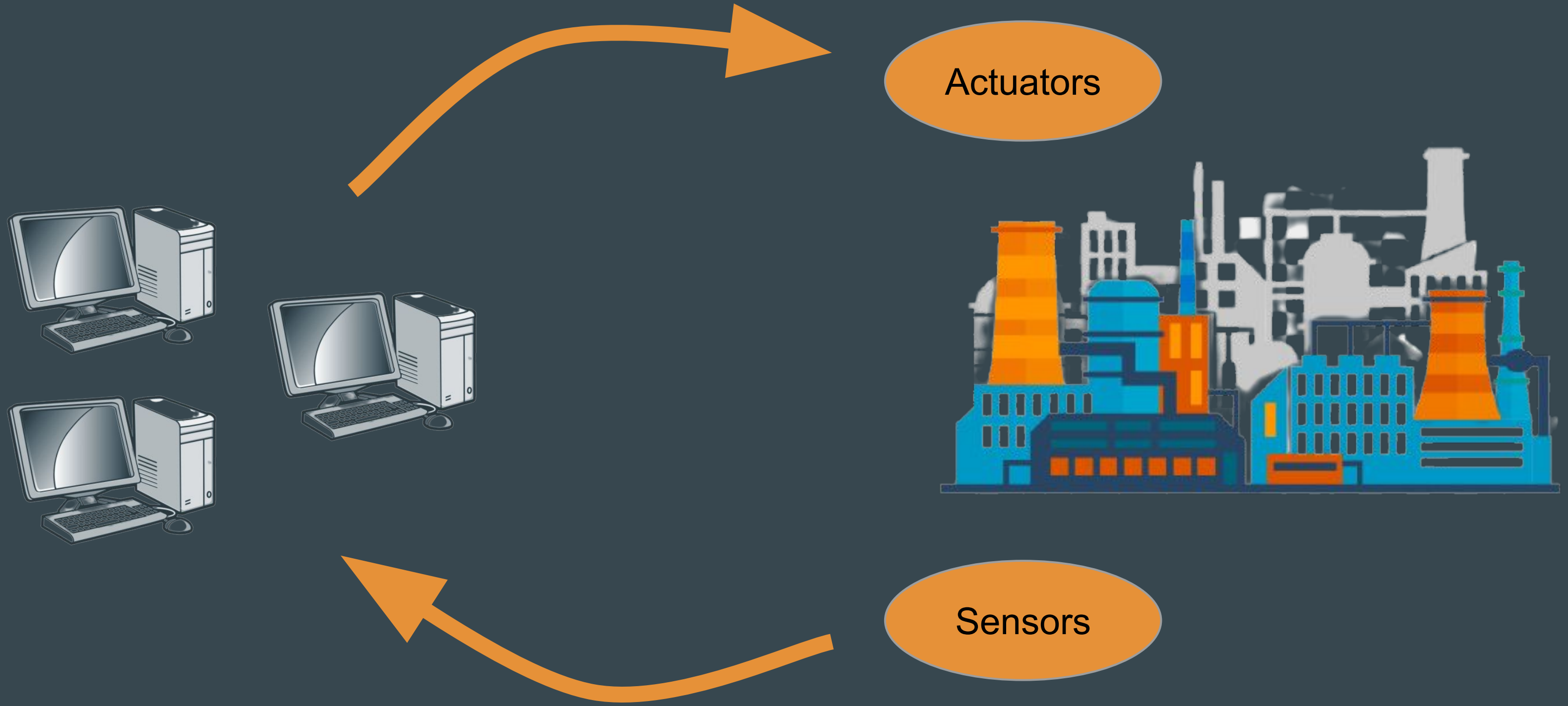
Prakhar Ganesh[1], Xin Lou[1], Yao Chen[1], Rui Tan[2], David K.Y. Yau[3], Deming Chen[4], Marianne Winslett[4]

[1] Advanced Digital Sciences Center, Illinois at Singapore
[2] Nanyang Technological University, Singapore
[3] Singapore University of Technology and Design
[4] University of Illinois at Urbana-Champaign, USA

Background

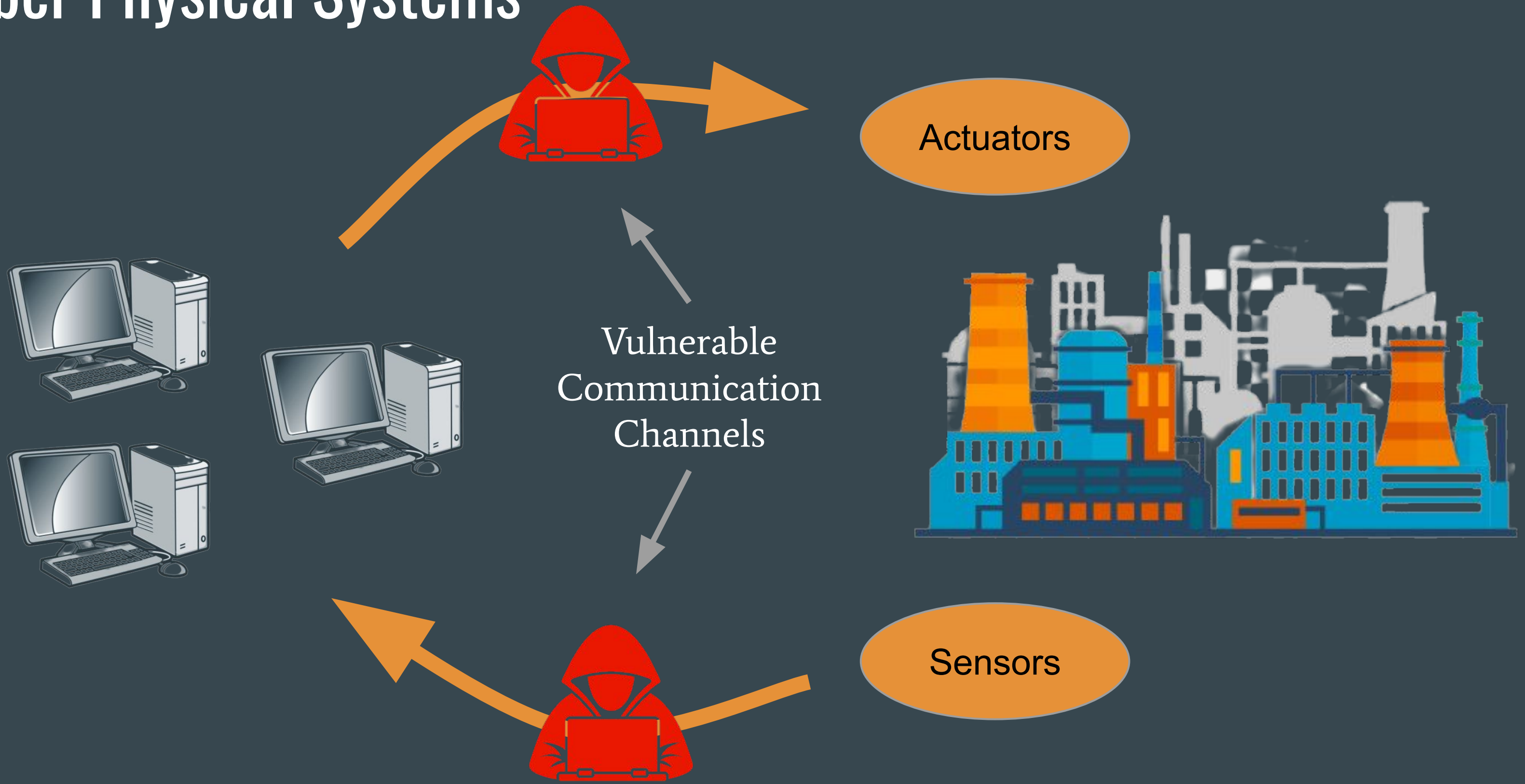
- Cyber-Physical Systems
- Time Delay Attack
- System and Threat Models

Cyber Physical Systems



H. Farhangi, “The path of the smart grid”, IEEE power and energy magazine, vol. 8, no. 1, pp. 18–28, 2009.

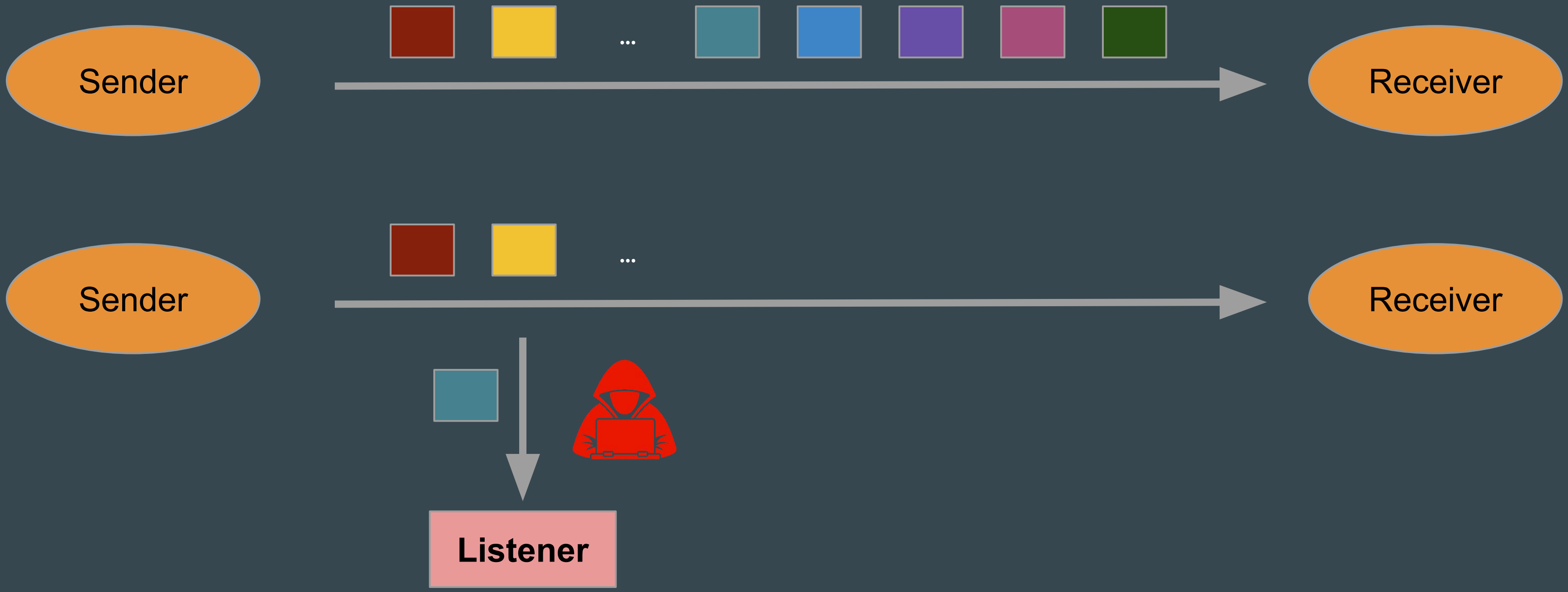
Cyber Physical Systems



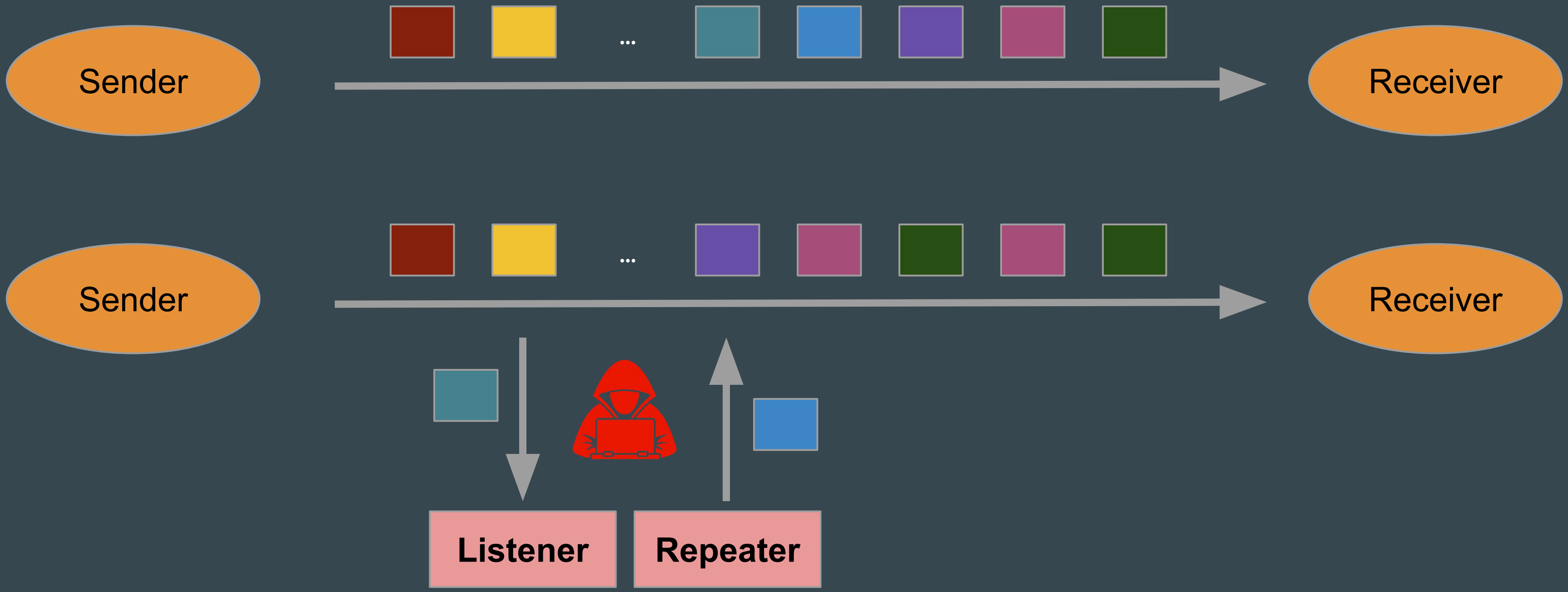
H. Farhangi, “The path of the smart grid”, IEEE power and energy magazine, vol. 8, no. 1, pp. 18–28, 2009.

Time Delay Attack

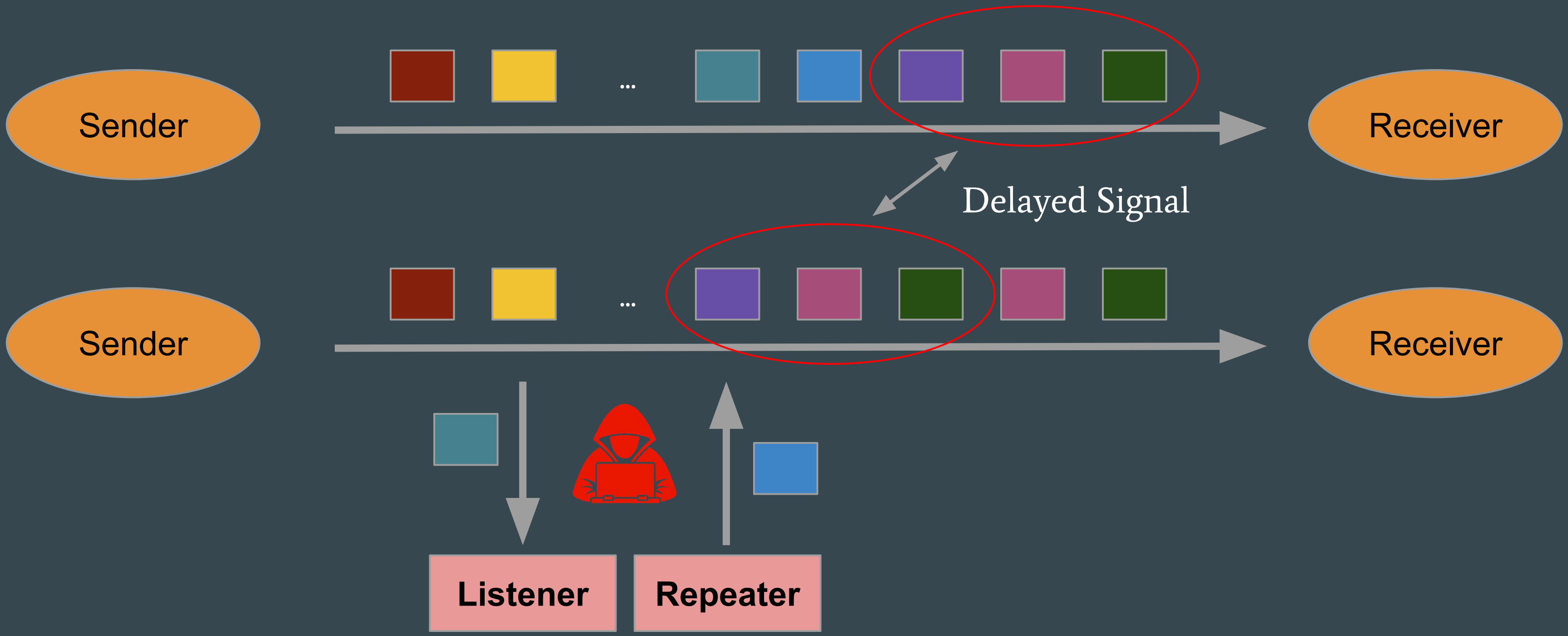
Time Delay Attack



Time Delay Attack

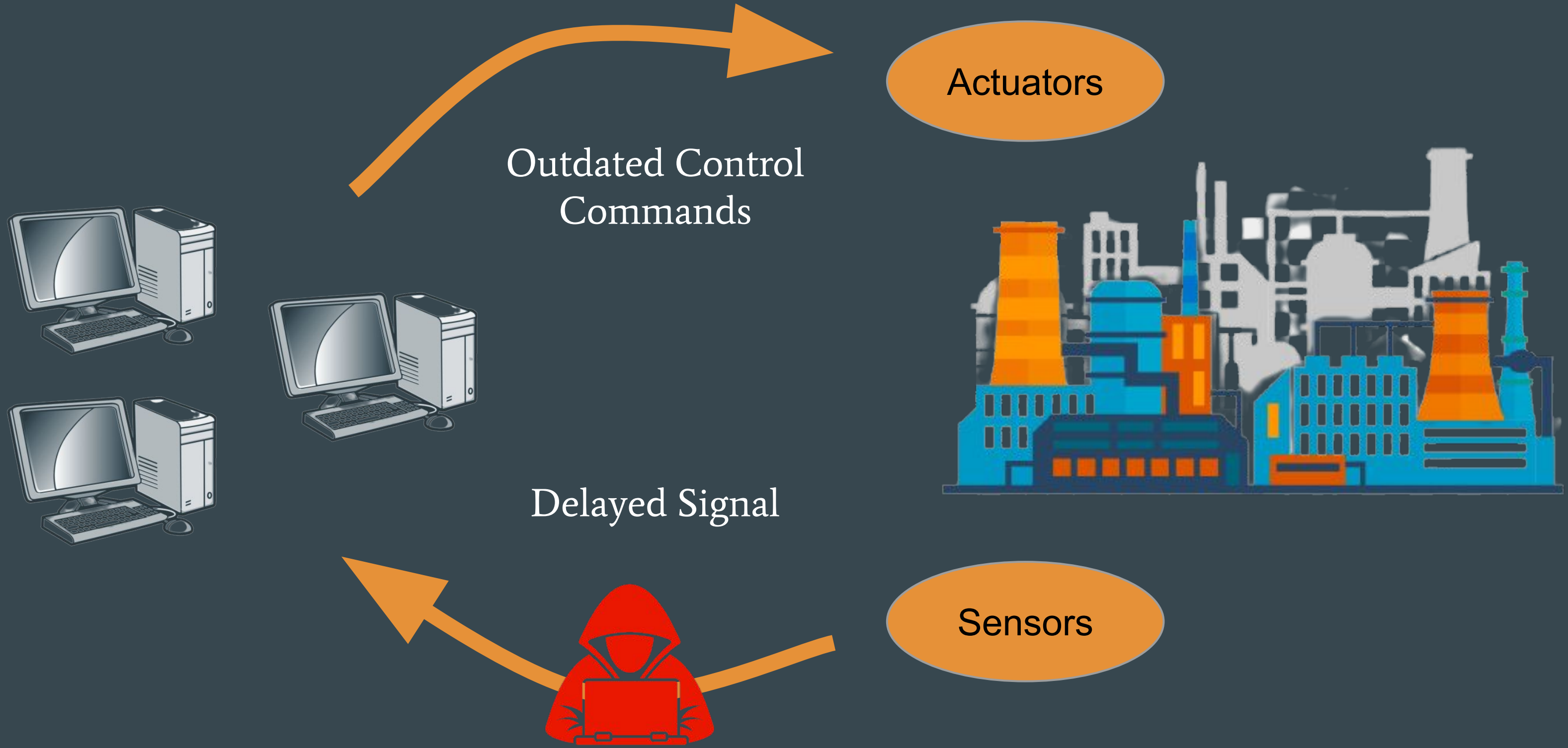


Time Delay Attack



A. Sargolzaei, et al., “Time-delay switch attack on load frequency control in smart grid”, Advances in Communication Technology, vol. 5, pp. 55–64, 2013.

Cyber Physical Systems

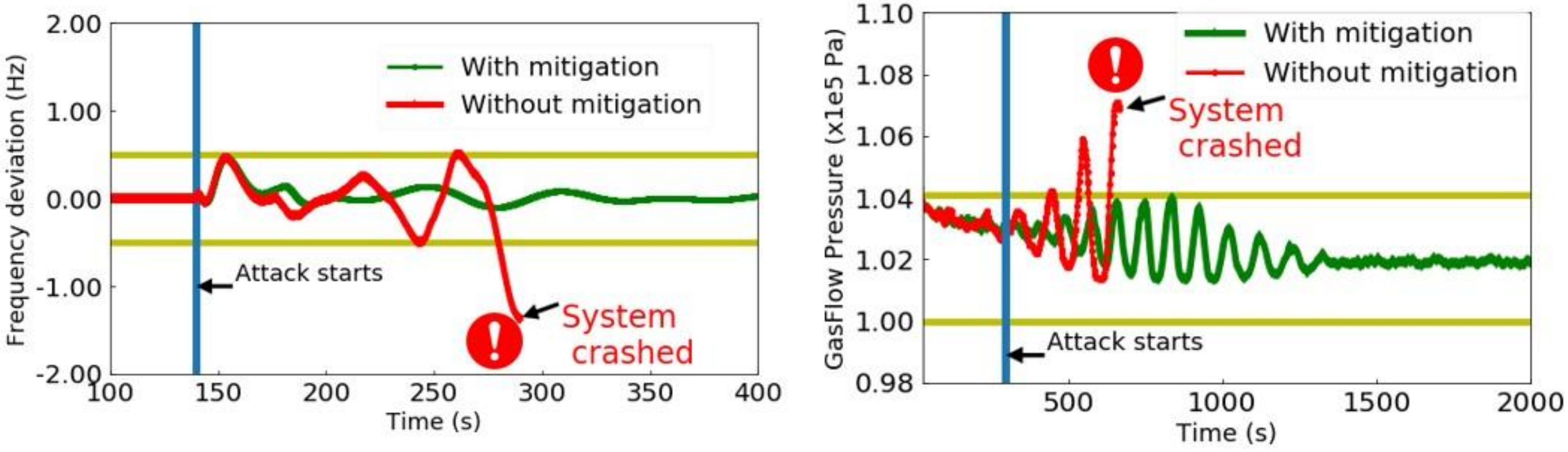


Why TDA?

- No abrupt change in signal traffic
- Does not require breaking the signal encryptions
- Can cause serious damage in closed loop systems

A. Sargolzaei, et al., “Time-delay switch attack on load frequency control in smart grid”, Advances in Communication Technology, vol. 5, pp. 55–64, 2013.

Impact Assessment and Mitigation



(a) AGC output.

(b) Power plant output.

System Model

- Closed loop discrete-time CPS control systems (PPCS and AGC)
- Time is divided into slots
- Running simulations to create both training and testing dataset
- System is subjected to disturbances, e.g., measurement noises, actuation biases, setpoint changes, etc

X. Lou, et al., “Assessing and mitigating impact of time delay attack : Case study for power grid controls”, IEEE JSAC, vol. 38, no. 1, pp.141–155, 2020.

Threat Model

- Packet is maliciously delayed by τ but not tampered (τ is an integer)
- Attack launched with a random delay value (τ) and a random delay location
- Assumption : Lack of a trustworthy clock synchronization between the controller and the actuator

X. Lou, et al., “Assessing and mitigating impact of time delay attack : Case study for power grid controls”, IEEE JSAC, vol. 38, no. 1, pp.141–155, 2020.

Challenges & Motivation

- Model-driven vs Data-driven methods
- Real-Time vs Post-mortem analysis
- Long Input Streams

Model-driven vs Data-driven methods

- Mathematical modeling creates highly complex models which are not robust to real-world noise
- Data-driven methods can learn to extract useful latent features that cannot be modelled manually
- Data-driven methods are easier to generalize and does not require domain expertise

Real-time vs Post-mortem analysis

<u>Real-time Analysis</u>

- Can help prevent damage
- Input information is a continuous data stream
- Can be inaccurate initially, but has the ability to improve over time

<u>Post-mortem Analysis</u>

- Does not have any direct practical use
- Complete input trace is available.
- Can be fairly accurate as it has seen the complete input signal

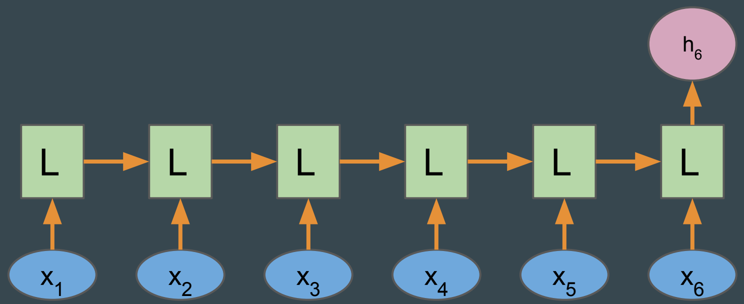
Long continuously running input stream

- Updating the output based on only the new input
- Processing the complete input signal at every time step can be very expensive
- Long input streams can lose information from the past. Traditional LSTMs are not suitable for the task

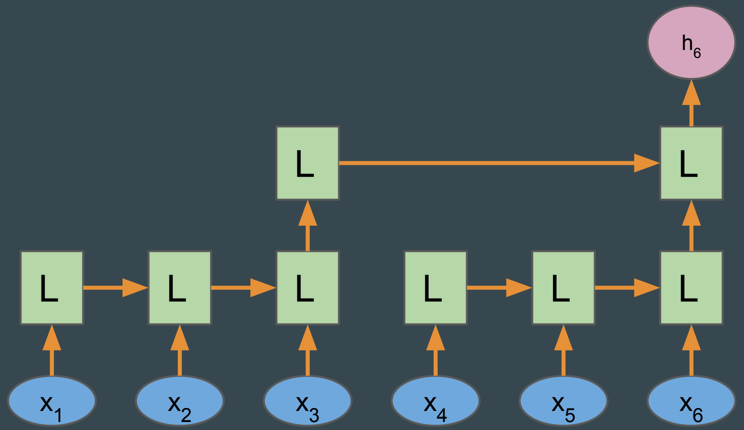
Our Solution

- Hierarchical LSTM
- Multi-head Output
- Asynchronous Training
- Interpretation Strategies
- Evaluation Results

Hierarchical LSTM

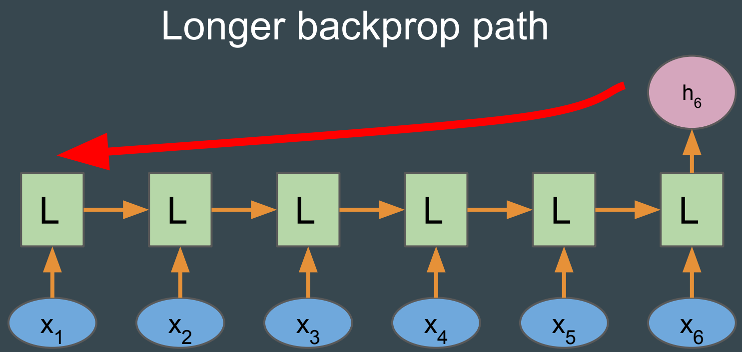


Traditional LSTM

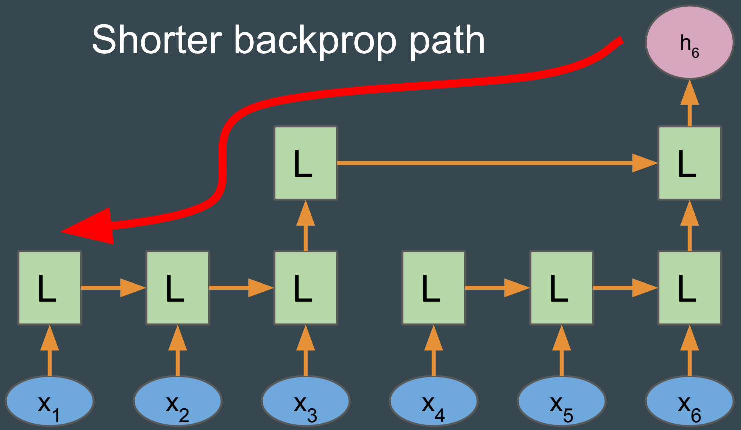


Hierarchical LSTM

Hierarchical LSTM

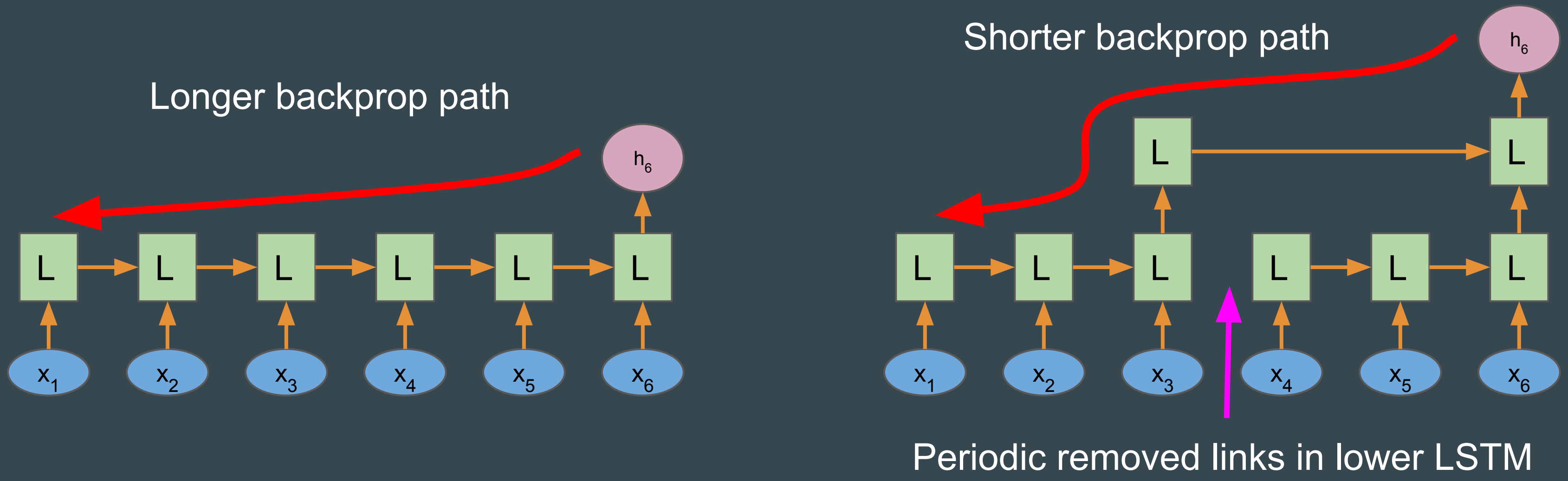


Traditional LSTM



Hierarchical LSTM

Hierarchical LSTM

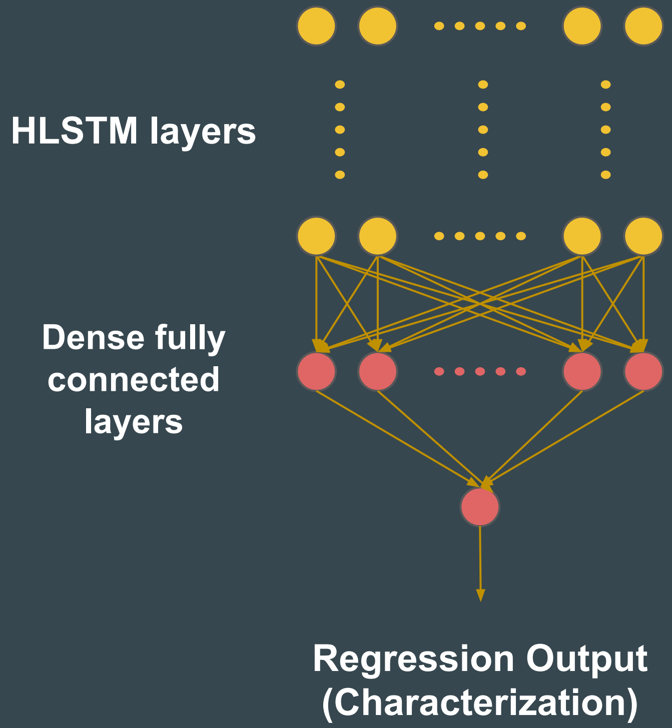


Traditional LSTM

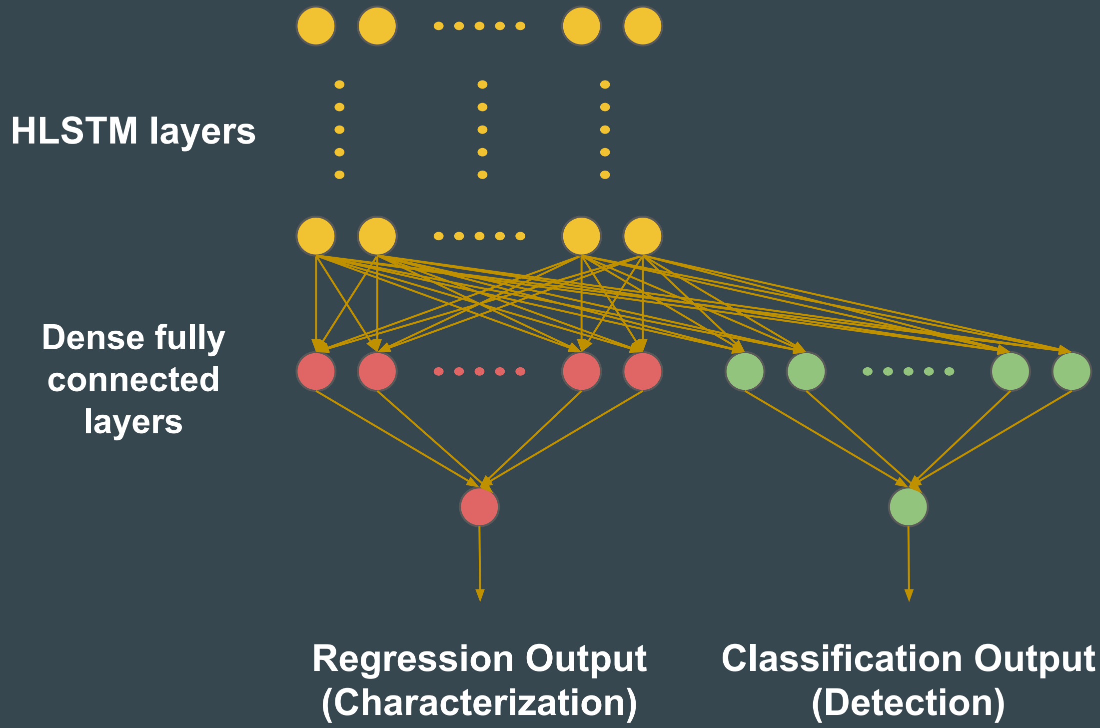
Hierarchical LSTM

Specialized Detection and Characterization

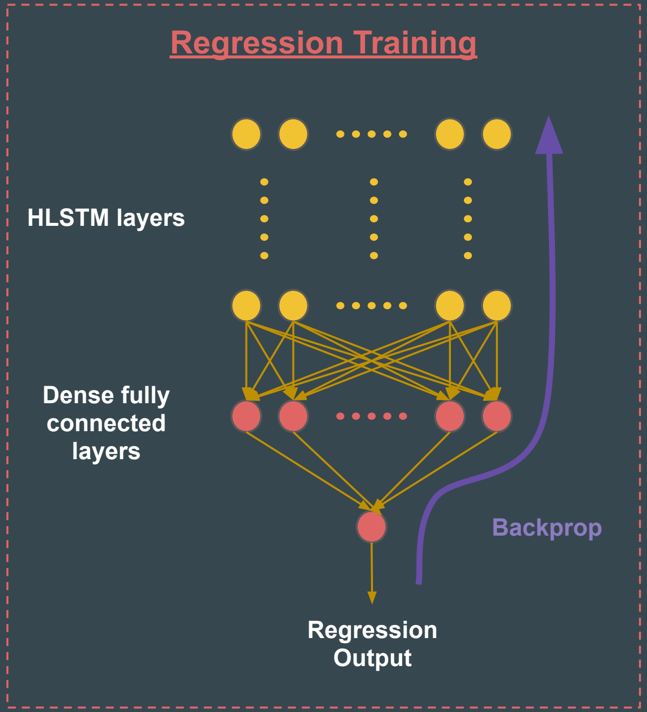
Specialized Detection and Characterization



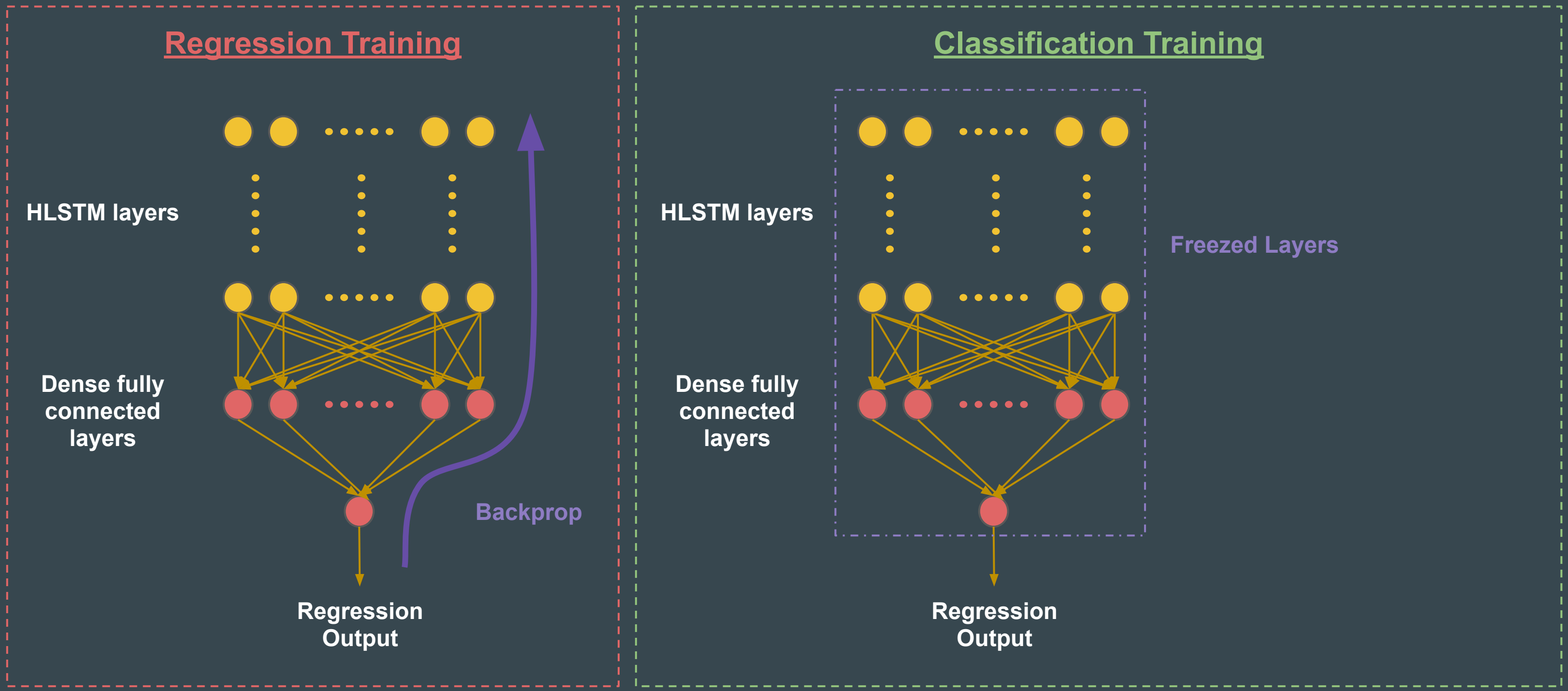
Specialized Detection and Characterization



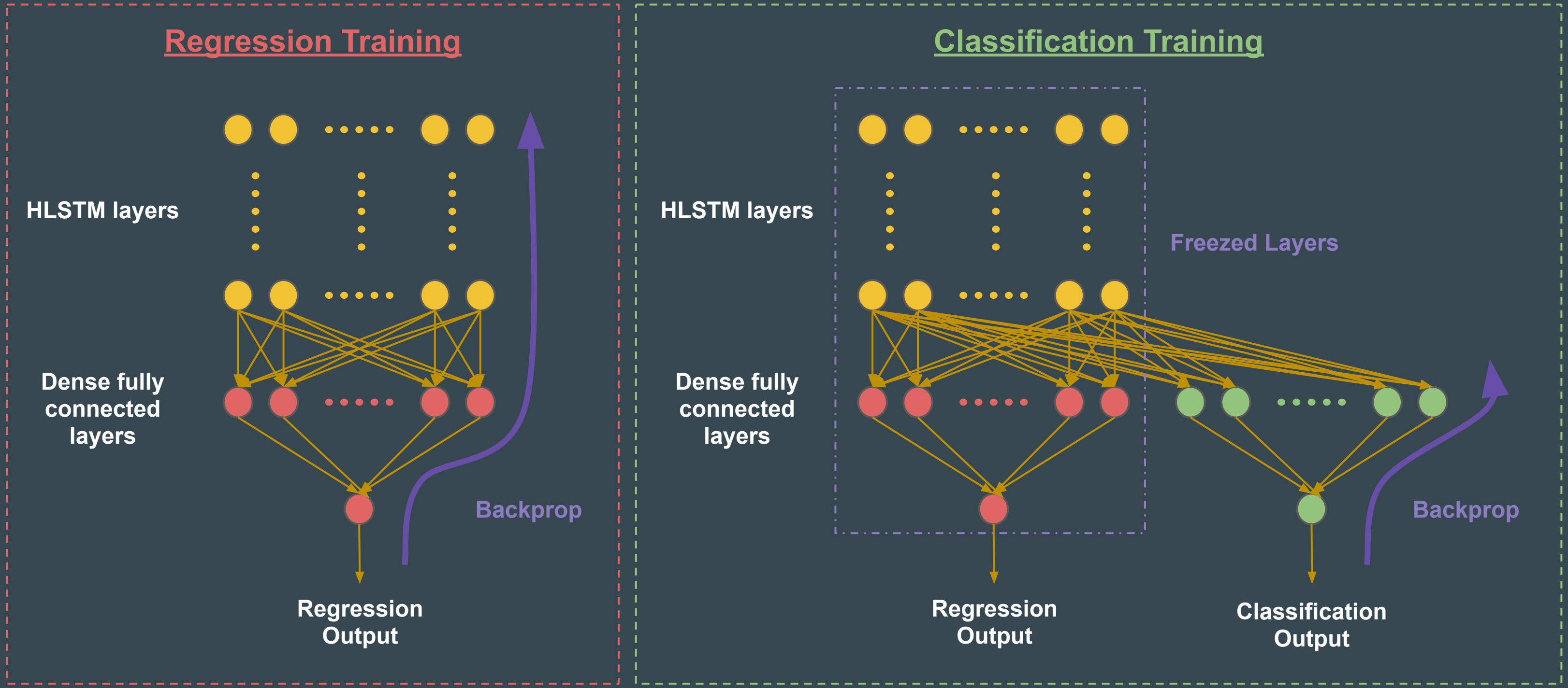
Asynchronous Training



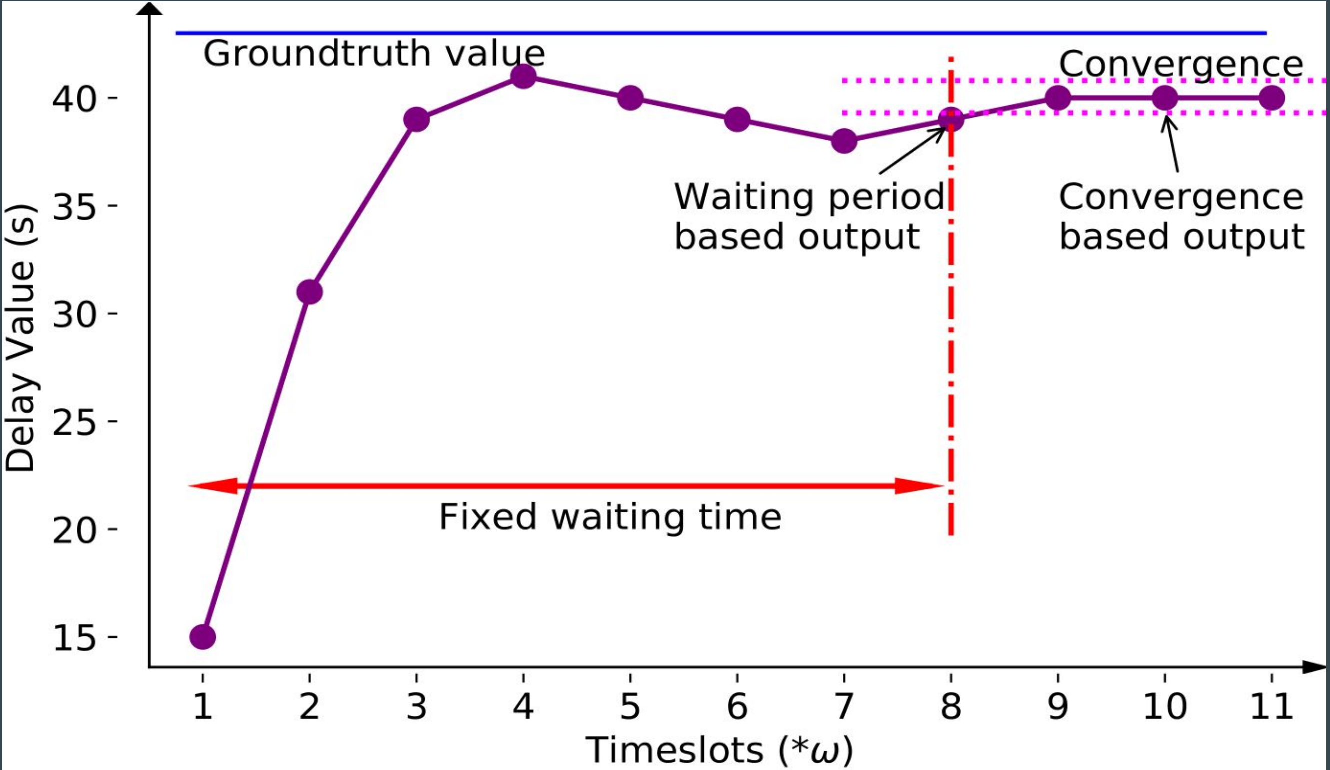
Asynchronous Training



Asynchronous Training



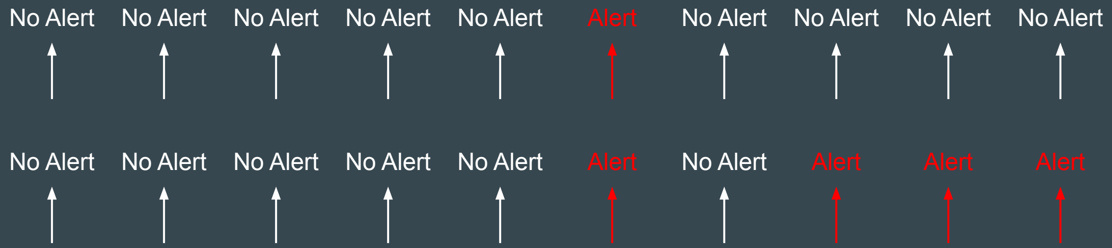
Flexible Interpretability (Regression)



Flexible Interpretability (Classification)

No Alert No Alert No Alert No Alert No Alert Alert No Alert No Alert No Alert No Alert

Flexible Interpretability (Classification)



Flexible Interpretability (Classification)

No Alert	No Alert	No Alert	No Alert	No Alert	Alert	No Alert	No Alert	No Alert	No Alert
No Alert	No Alert	No Alert	No Alert	No Alert	Alert	No Alert	Alert	Alert	Alert
No Alert	No Alert	No Alert	No Alert	No Alert	Alert	Alert	Alert	Alert	Alert

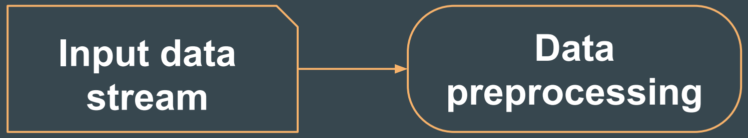
Flexible Interpretability (Classification)

No Alert	No Alert	No Alert	No Alert	No Alert	Alert	No Alert	No Alert	No Alert	No Alert
No Alert	No Alert	No Alert	No Alert	No Alert	Alert	No Alert	Alert	Alert	Alert
No Alert	No Alert	No Alert	No Alert	No Alert	Alert	Alert	Alert	Alert	Alert
No Alert	No Alert	No Alert	No Alert	No Alert	Alert	Alert	No Alert	No Alert	No Alert

Complete Model

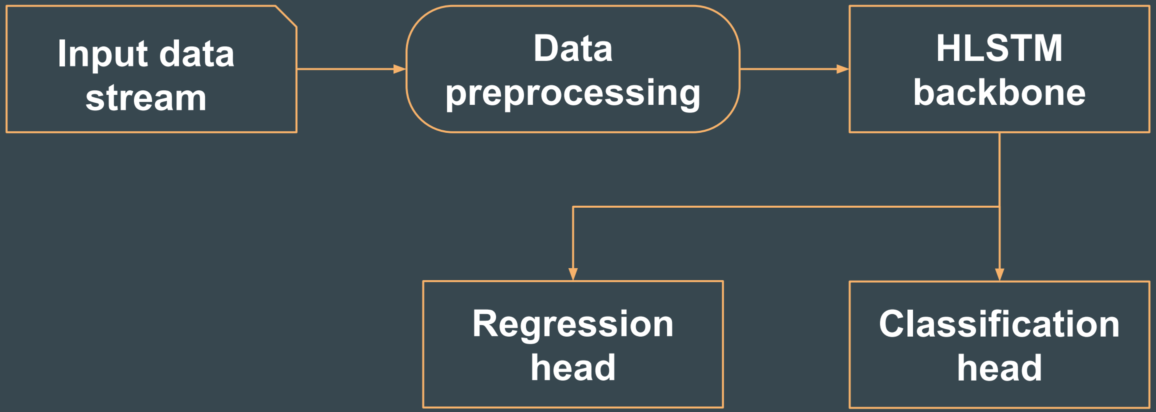
Input data stream

Complete Model

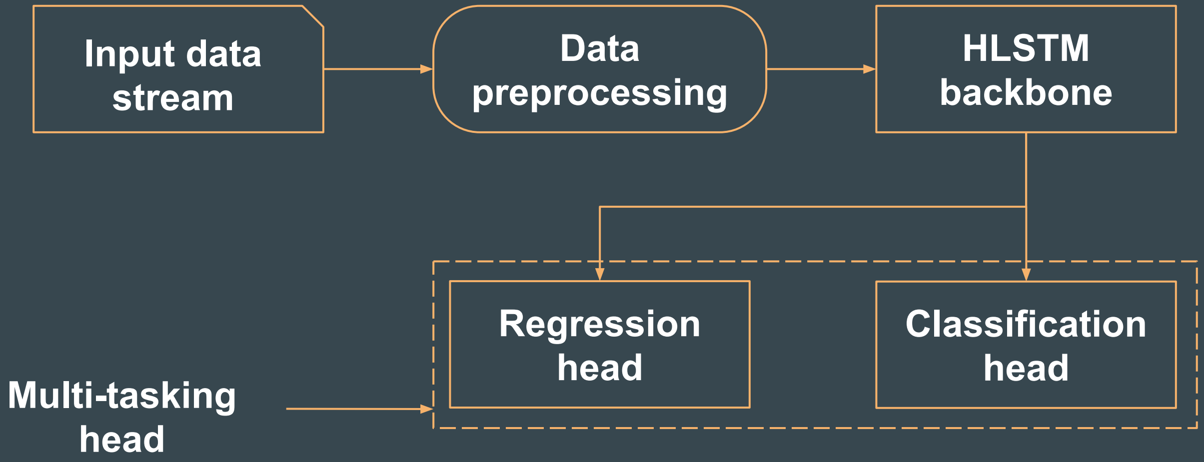


Complete Model

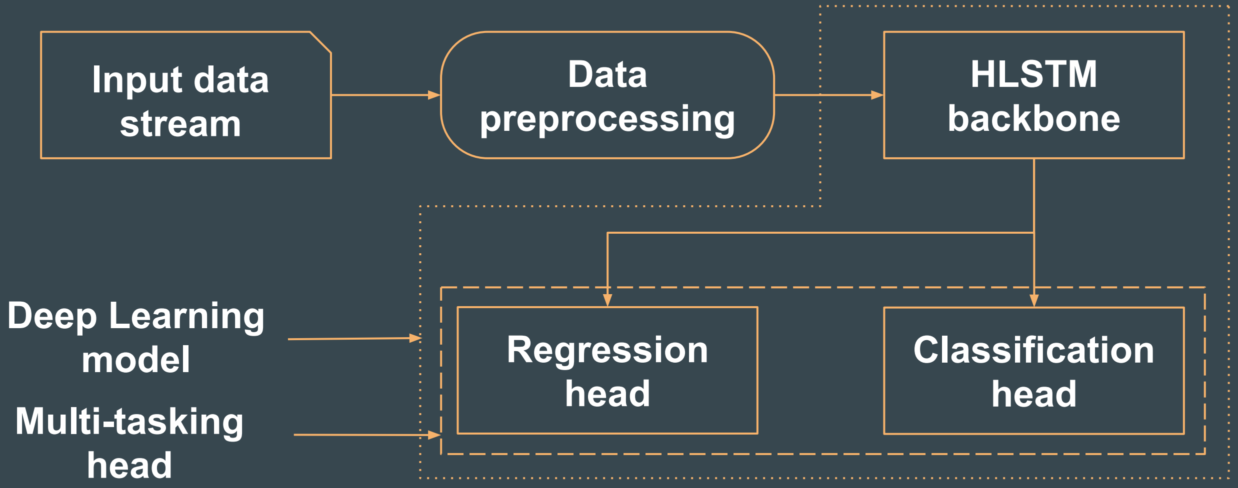
Complete Model



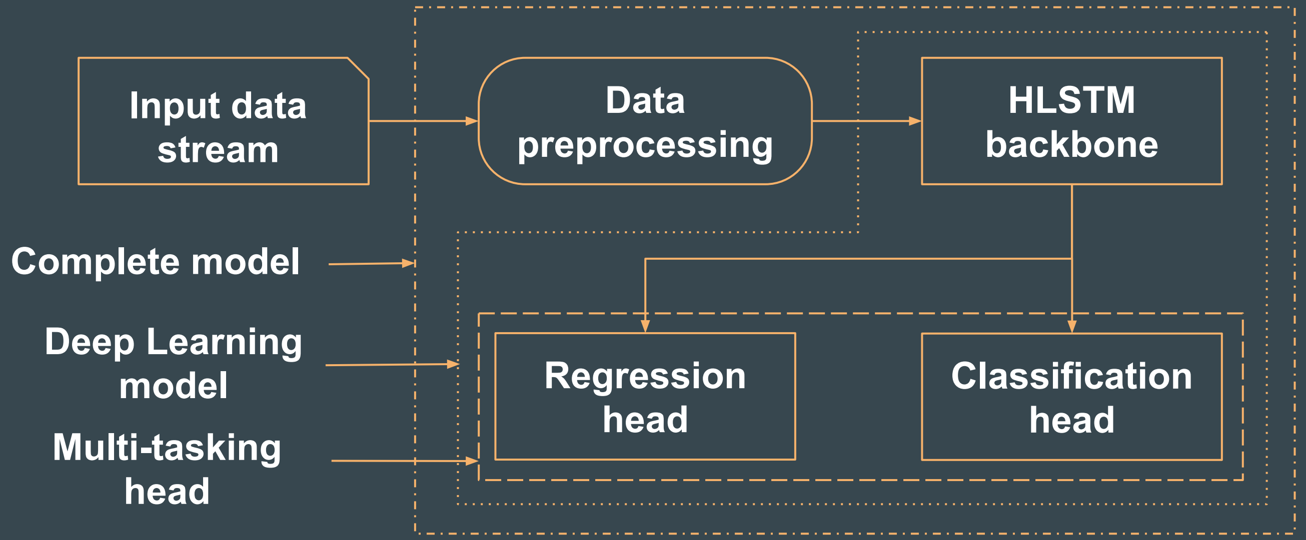
Complete Model



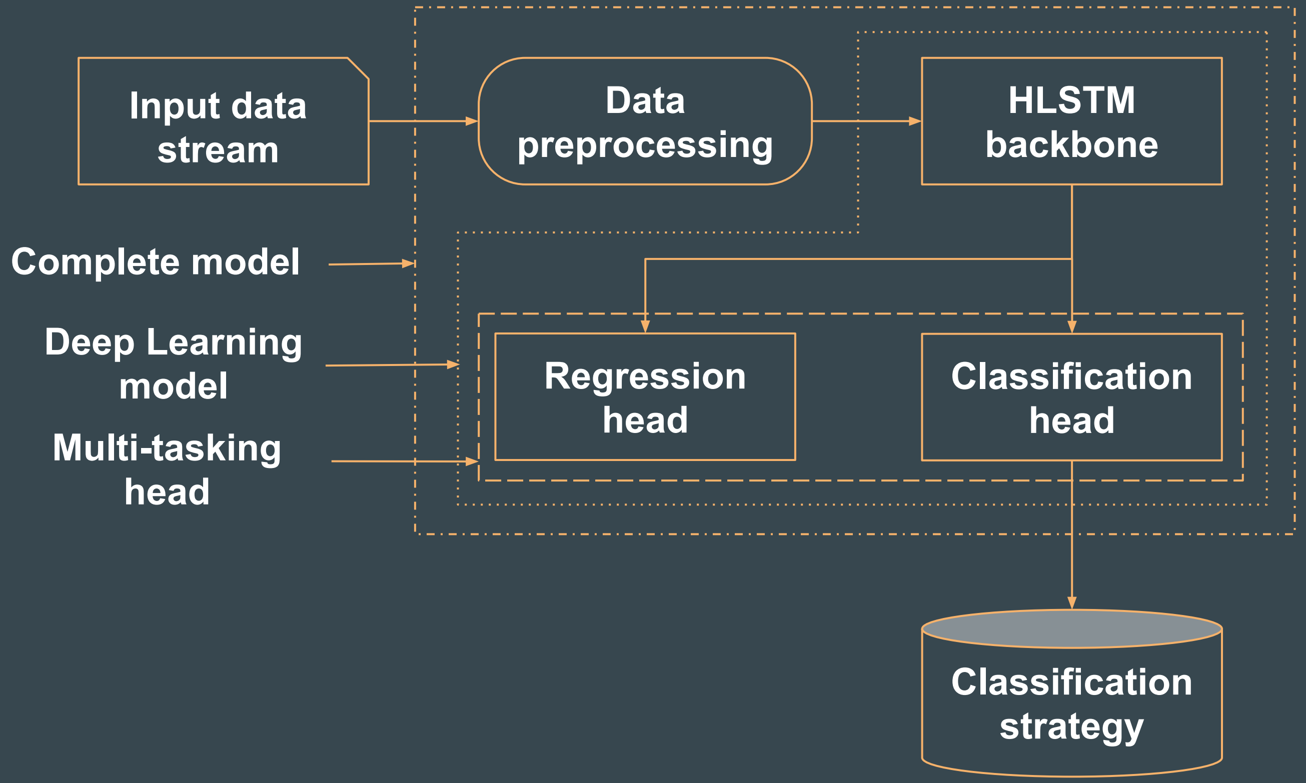
Complete Model



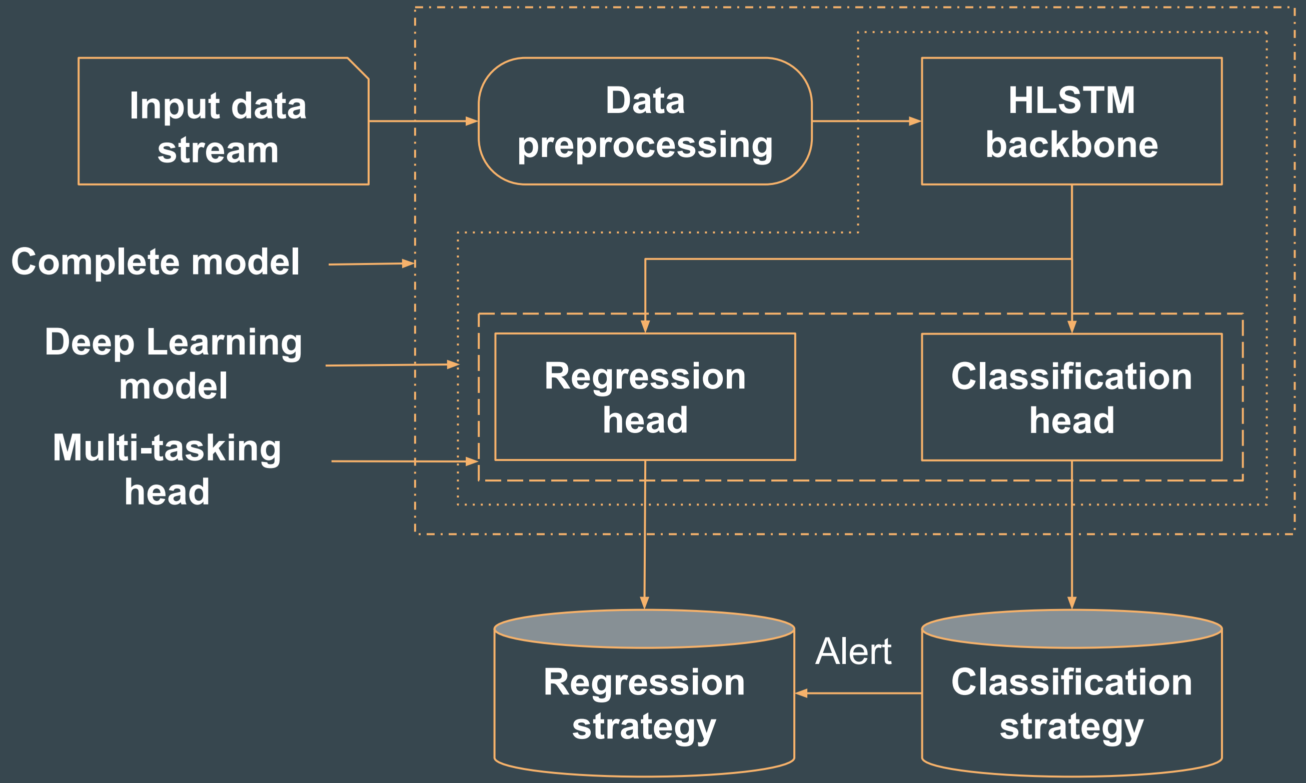
Complete Model



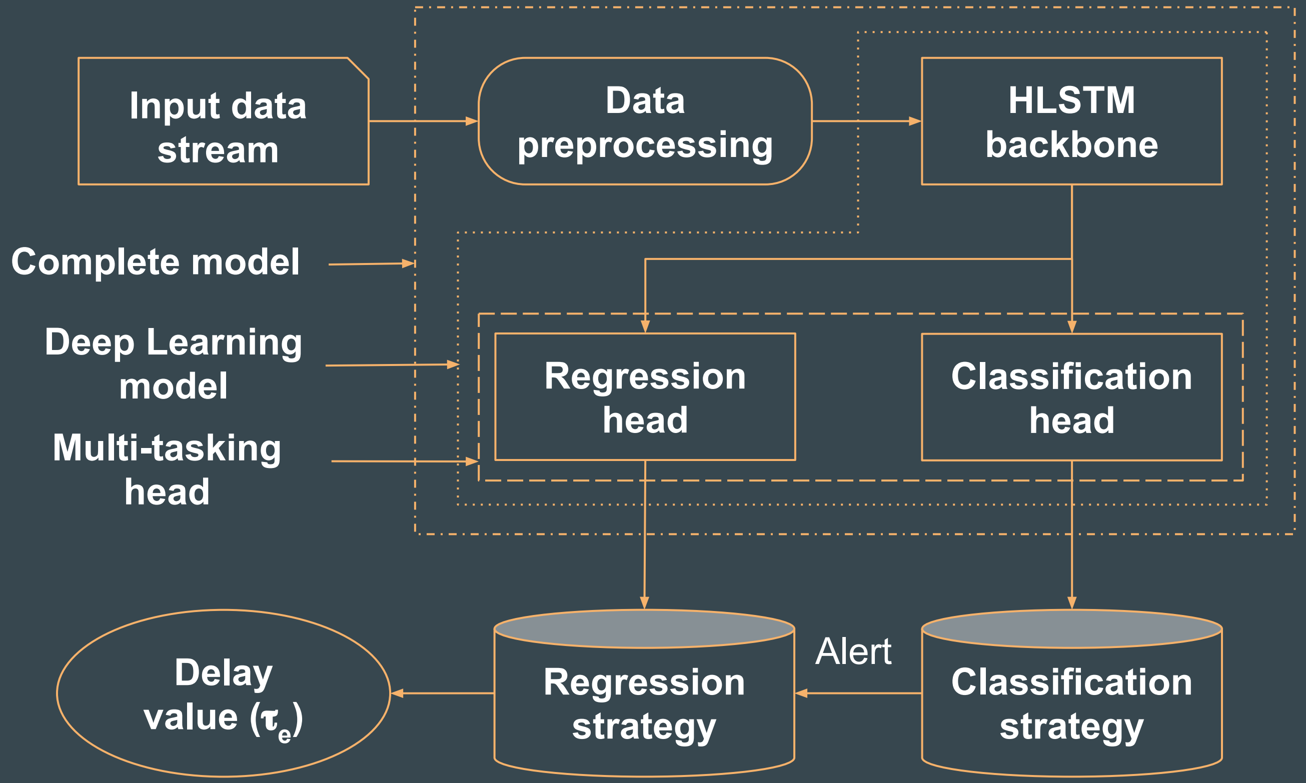
Complete Model



Complete Model



Complete Model



Evaluation Metrics

- MAE and RMSE

  -- To evaluate the characterization head
- Accuracy, False Positive and False Negative

  -- To evaluate the detection head
- T_{avg}

  -- To evaluate average characterization latency

Internal Ablation Study (PPCS)

- HLSTM outperforms Accumulate LSTM or Vanilla LSTM, under the same multi-task head training and testing setup

Approach	Classification (Detection)			Regression (Characterization)		
	Accuracy	FP	FN	MAE	RMSE	T_{avg}
Vanilla LSTM + Multi-task	85.21%	9.7%	5.1%	4.17	8.76	136
Accumulate LSTM + Multi-task	89.76%	7.1%	3.3%	2.76	6.03	168
HLSTM + Multi-task	92.39%	4.7%	2.9%	2.03	5.48	128

Internal Ablation Study (PPCS)

- HLSTM outperforms Accumulate LSTM or Vanilla LSTM, under the same multi-task head training and testing setup
- Regression only method is worse without a specialised classification head

Approach	Classification (Detection)			Regression (Characterization)		
	Accuracy	FP	FN	MAE	RMSE	T_{avg}
Vanilla LSTM + Multi-task	85.21%	9.7%	5.1%	4.17	8.76	136
Accumulate LSTM + Multi-task	89.76%	7.1%	3.3%	2.76	6.03	168
HLSTM + Regression only	--	--	--	3.51	7.46	148
HLSTM + Multi-task	92.39%	4.7%	2.9%	2.03	5.48	128

Internal Ablation Study (PPCS)

- HLSTM outperforms Accumulate LSTM or Vanilla LSTM, under the same multi-task head training and testing setup
- Regression only method is worse without a specialised classification head
- 2 separate HLSTM backbones for detection and characterization perform slightly better than a common backbone, but doubles the number of computations

Approach	Classification (Detection)			Regression (Characterization)		
	Accuracy	FP	FN	MAE	RMSE	T_{avg}
Vanilla LSTM + Multi-task	85.21%	9.7%	5.1%	4.17	8.76	136
Accumulate LSTM + Multi-task	89.76%	7.1%	3.3%	2.76	6.03	168
HLSTM + Regression only	--	--	--	3.51	7.46	148
HLSTM + Multi-task	92.39%	4.7%	2.9%	2.03	5.48	128
2 HLSTM + Multi-task	93.03%	3.7%	3.2%	2.02	5.53	124

Internal Ablation Study (PPCS)

- HLSTM outperforms Accumulate LSTM or Vanilla LSTM, under the same multi-task head training and testing setup
- Regression only method is worse without a specialised classification head
- 2 separate HLSTM backbones for detection and characterization perform slightly better than a common backbone, but doubles the number of computations

Approach	Classification (Detection)			Regression (Characterization)		
	Accuracy	FP	FN	MAE	RMSE	T_{avg}
Vanilla LSTM + Multi-task	85.21%	9.7%	5.1%	4.17	8.76	136
Accumulate LSTM + Multi-task	89.76%	7.1%	3.3%	2.76	6.03	168
HLSTM + Regression only	--	--	--	3.51	7.46	148
HLSTM + Multi-task	92.39%	4.7%	2.9%	2.03	5.48	128
2 HLSTM + Multi-task	93.03%	3.7%	3.2%	2.02	5.53	124

Comparing Against Baseline Models (PPCS)

- Deep learning models can reduce the error to almost half, when compared against traditional models like kNN and RF
- All existing methods provide post-mortem analysis, but our method can provide real-time results, reducing the average reaction latency from 300s to 128s while improving the error even further

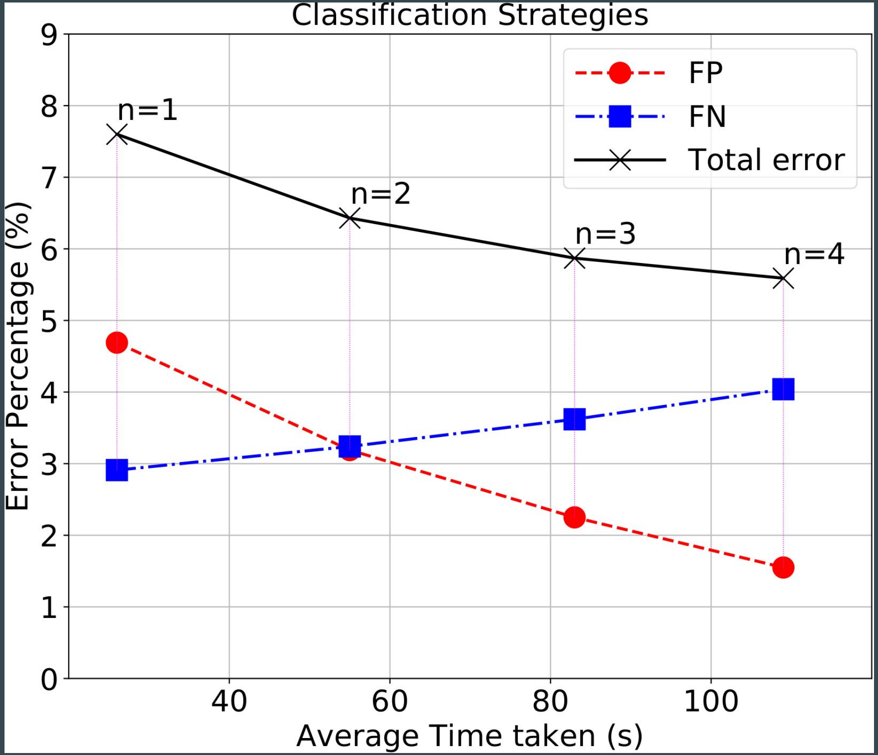
Approach	Classification (Detection)			Regression (Characterization)		
	Accuracy	FP	FN	MAE	RMSE	T_{avg}
kNN	72.6%	11.8%	15.6%	6.23	9.48	300
Random Forest	80.82%	5.2%	13.9%	6.44	10.32	300
(Lou et al, 2019)	--	--	--	3.73	6.84	300
Our Model	92.39%	4.7%	2.9%	2.03	5.48	128

X. Lou, et al., “Learning-based time delay attack characterization for cyber-physical systems”, IEEE SmartGridComm, AI in Energy Systems (Invited Paper), 2019.

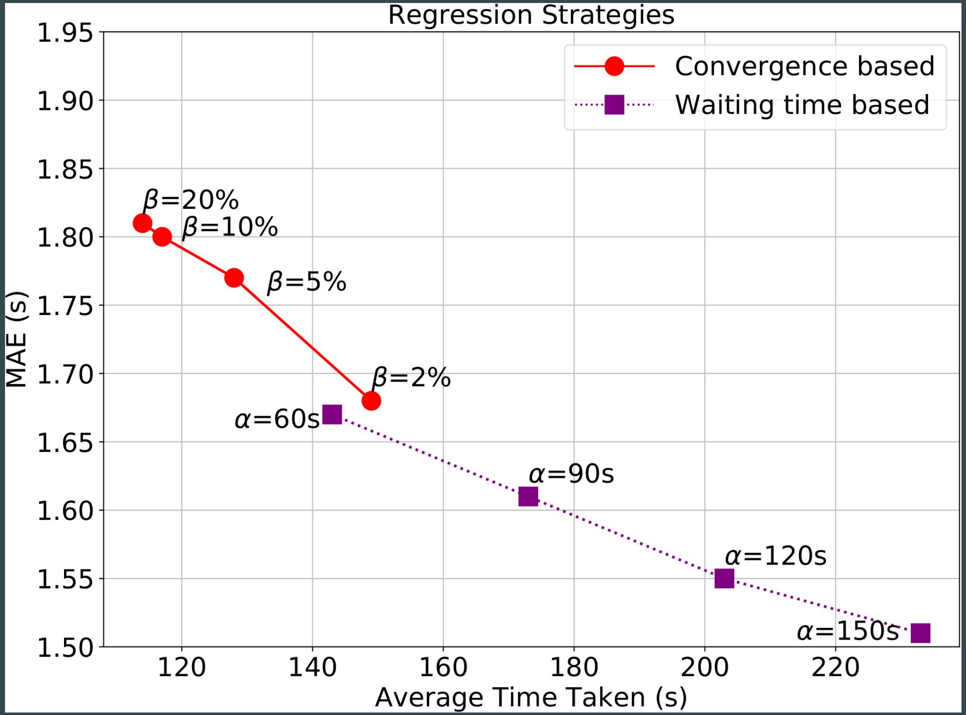
Comparing Against Baseline Models (AGC)

Approach	Classification (Detection)			Regression (Characterization)		
	Accuracy	FP	FN	MAE	RMSE	T_{avg}
kNN	71.29%	--	28.7%	3.27	5.02	200
Random Forest	74.57%	--	25.4%	3.41	4.10	200
(Lou et al, 2019)	--	--	--	1.34	2.17	200
Vanilla LSTM + Multi-task	81.07%	--	18.9%	1.74	3.84	63
Accumulate LSTM + Multi-task	97.71%	--	2.3%	0.77	1.42	65
HLSTM + Regression only	--	--	--	1.07	1.84	81
2 HLSTM + Multi-task	99.09%	--	0.9%	0.47	0.91	57
HLSTM + Multi-task	98.49%	--	1.5%	0.49	0.98	49

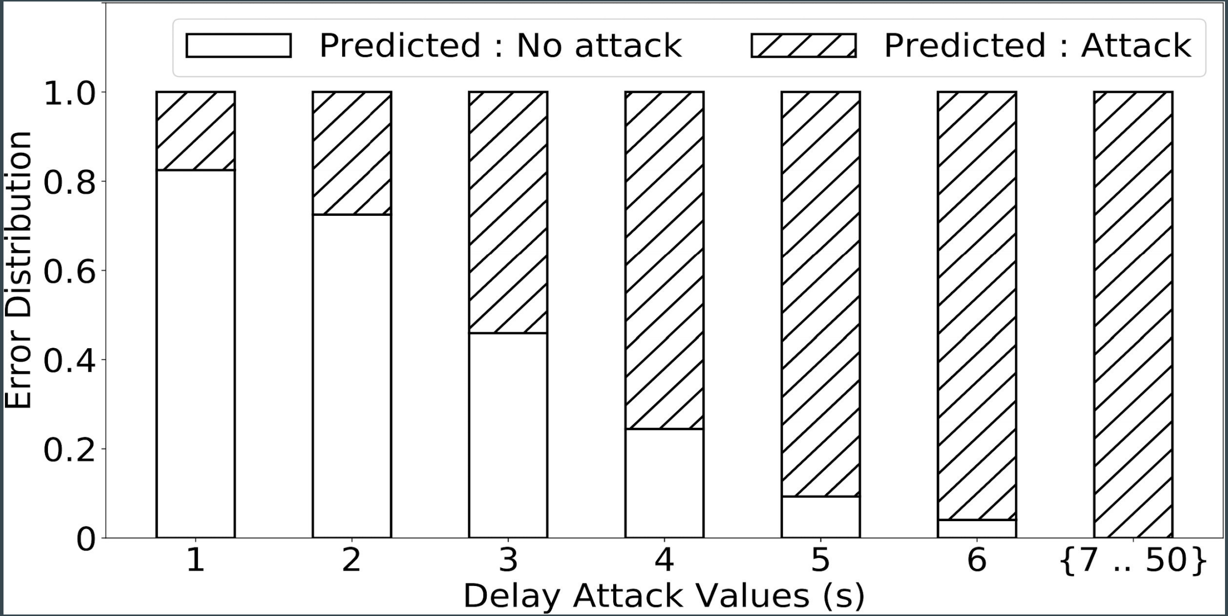
Interpretation Strategies (Classification)



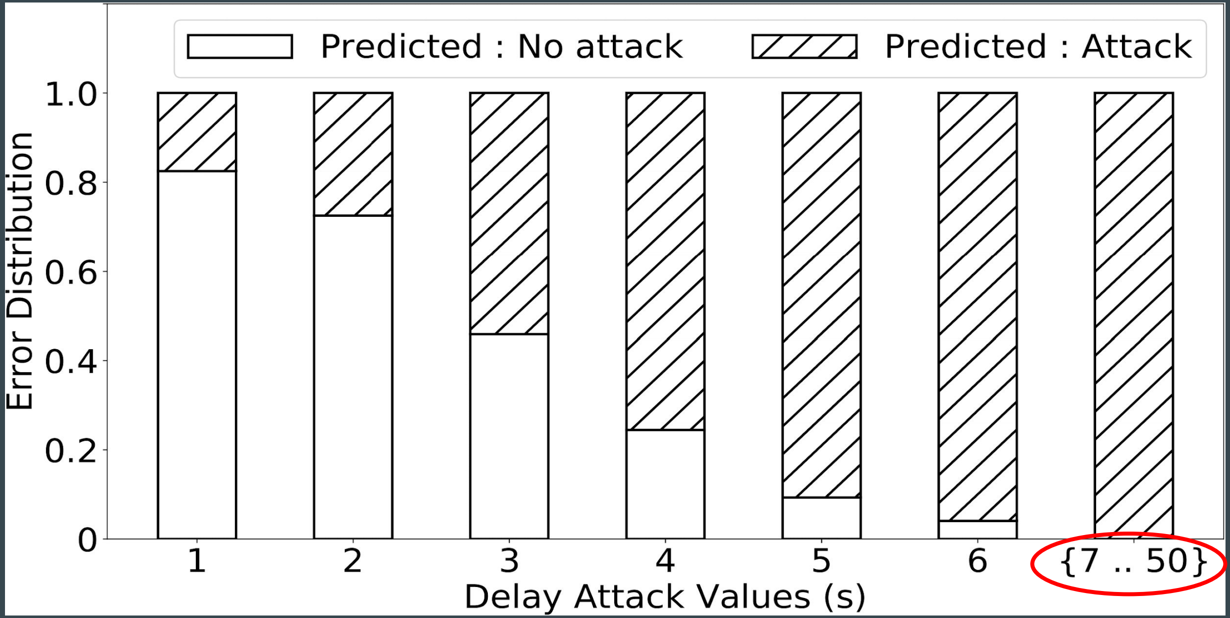
Interpretation Strategies (Regression)



Exploring Model Sensitivity (PPCS)



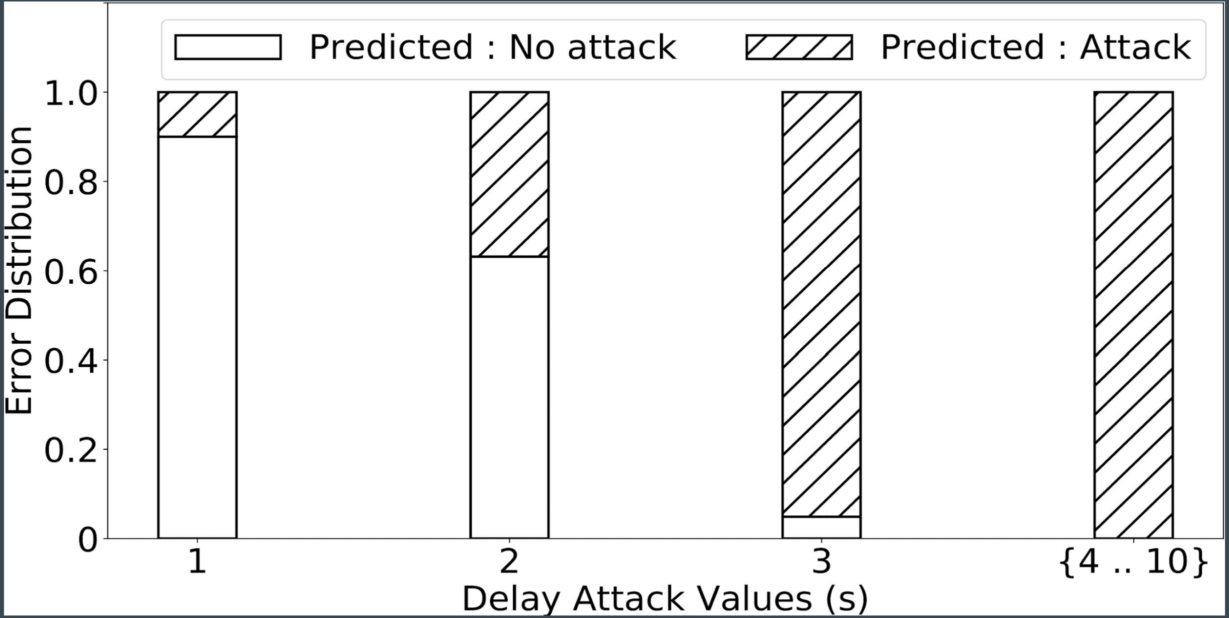
Exploring Model Sensitivity (PPCS)



PPCS can actually tolerate upto τ=12 value delay attack with no harm

X. Lou, et al., "Assessing and mitigating impact of time delay attack : Case study for power grid controls", IEEE JSAC, vol. 38, no. 1, pp.141–155, 2020.

Exploring Model Sensitivity (AGC)



In Conclusion

- Time Delay Attacks are unique cyber attacks that are difficult to detect and can cause real harm to the system
- A majority of existing solutions to TDA are dependent on mathematical modelling of the system, and perform post-mortem analysis.
- We propose a learning-based solution to detect and characterize TDA
- We propose hierarchical LSTM backbone to deal with longer sequences, multi-tasking head specialised to do detection and characterization, and various strategies to improve the interpretability of our model.
- We provide significant improvement in accuracy while reducing the reaction latency by 3-4 times.

Thank You