Deep Learning for Unsupervised Insider Threat Detection in Structured Cybersecurity Data Streams

Aaron Tuor

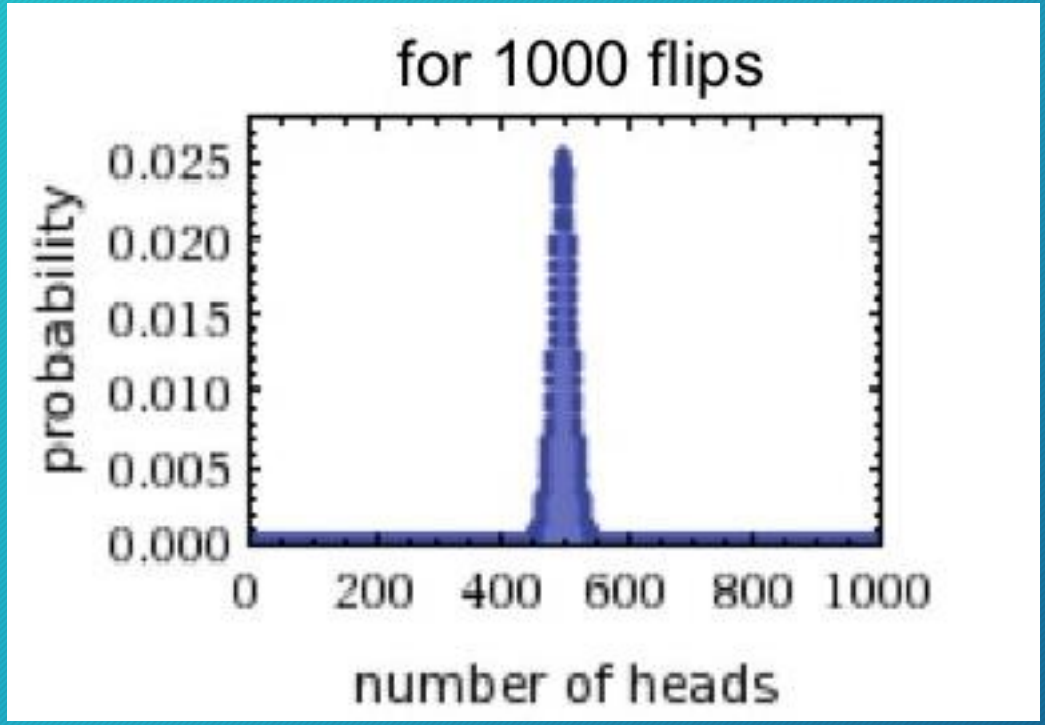
w/ Brian Hutchinson, Sam Kaplan, Nicole Nichols, Sean Robinson

Pacific Northwest National Laboratory

- Location in Richland and Seattle Washington
- Applications for internships are now open

  http://jobs.pnnl.gov/
- Spend a whole summer doing Computer Science research
- Collaborate with scientists from other fields
- Fun enrichment events:
  - Hanford museum
  - Lake Washington cruise
  - Virtual reality lab tour
  - Tour Laser Interferometer Gravitational-Wave Observatory (LIGO)

Anomaly Detection/ Outlier Analysis



Motivation

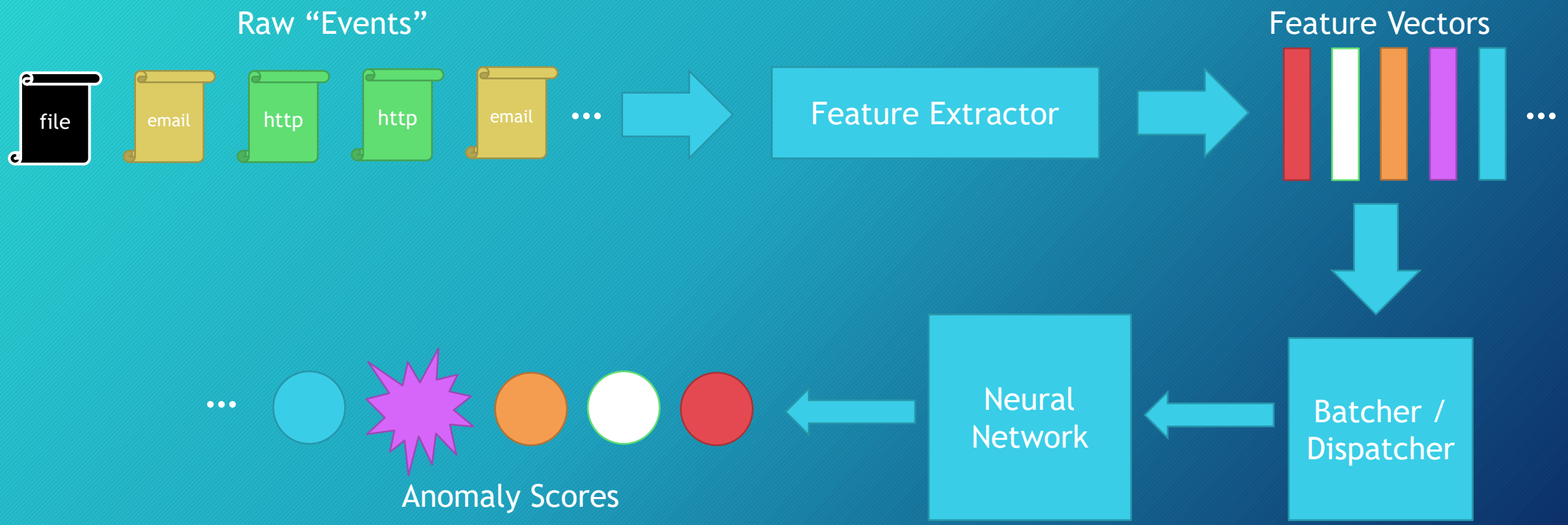
- Insider Threat: Actions taken by an employee harmful to an organization
  - Unsanctioned data transfer
  - Sabotage of resources
  - Misuse of network that disrupts organization
- Analysis of network activity can detect Insider Threat
- Automated filtering can help reduce the analyst workload

Approach Constraints

Approach should provide:

- Real time evaluation
- Upper bound on storage requirements
- Analysis of structured multivariate data
- Adaptation to shifting distribution of activities
- Interpretable assessments

System at a Glance



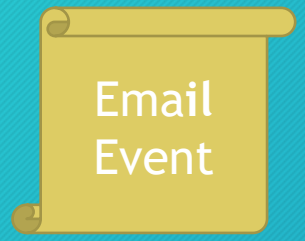
Outline

- **Data processing and feature extraction**
- Deep learning architectures
- Experiments and results
- Takeaways

Data Sources

- CERT Insider Threat (version 6.2)
  - ❖Synthetic data generated according to sophisticated user model
  - ❖516 days, 135 million events total
  - ❖Email, web, logon, file and device usage events
  - ❖5 bad actors produce 470 threat events
  - ❖Accompanying user meta data (role, project, team, …)

CERT: Example Log Line

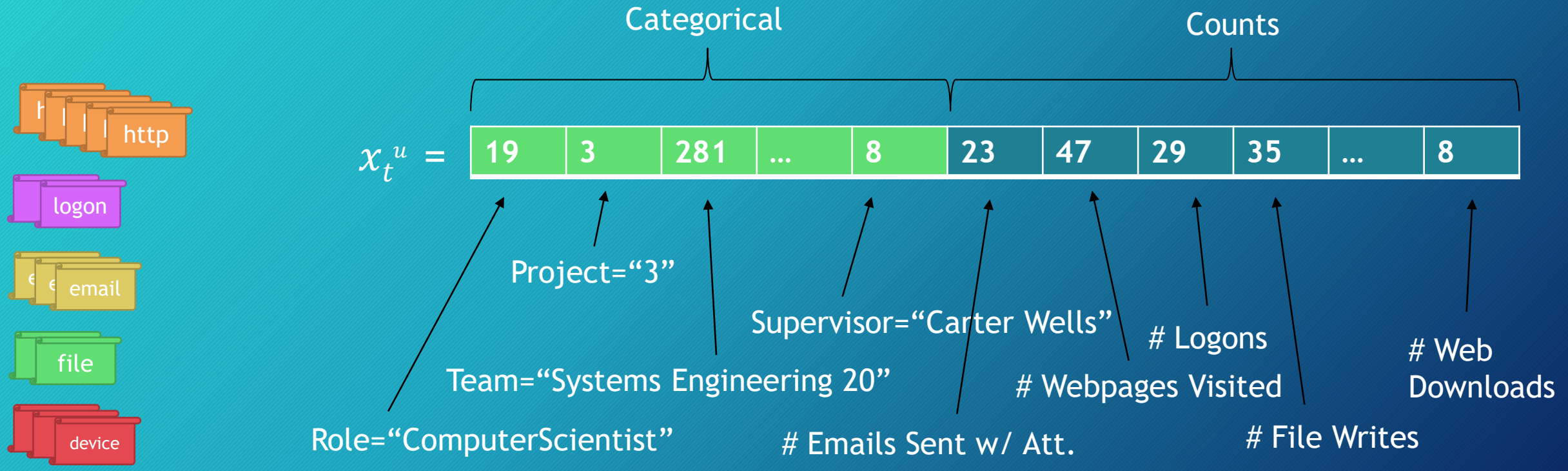


=

Event ID	I102-B4EB49RW-7379WSQW
Date	1/2/2010 6:36:41
User	HDB1666
PC	PC-6793
To	Louis.Bernard.Garza@dtaa.com
CC	Emery.Ali.Holloway@dtaa.com
BCC	Hector.Donovan.Bray@dtaa.com
From	Hector.Donovan.Bray@dtaa.com
Activity	Send
Size	45659
Attachments	<none>
Content	Now Sylvia, the object …

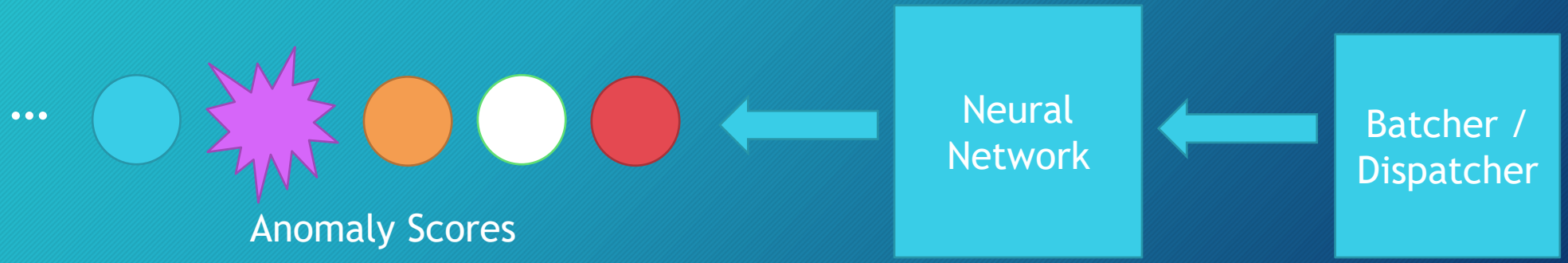
Aggregate Feature Vector

- For each user, aggregate their events over a window of time (e.g. one day)
- Example feature vector:

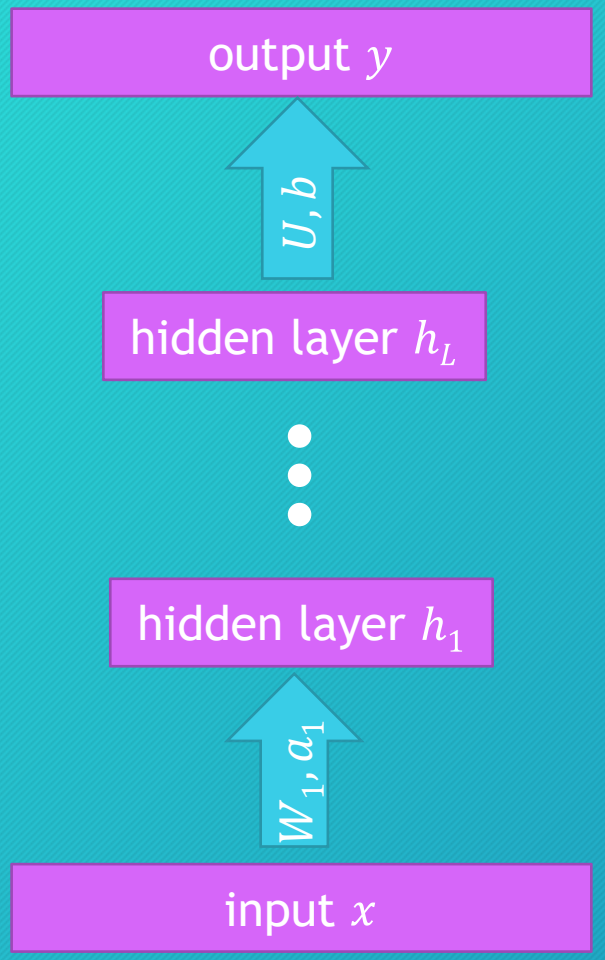


Outline

- Data processing and feature extraction
- **Deep learning architectures**
- Experiments and results
- Takeaways and ongoing work



Deep Neural Network Autoencoder



$$y = f(U^T h_L + b)$$

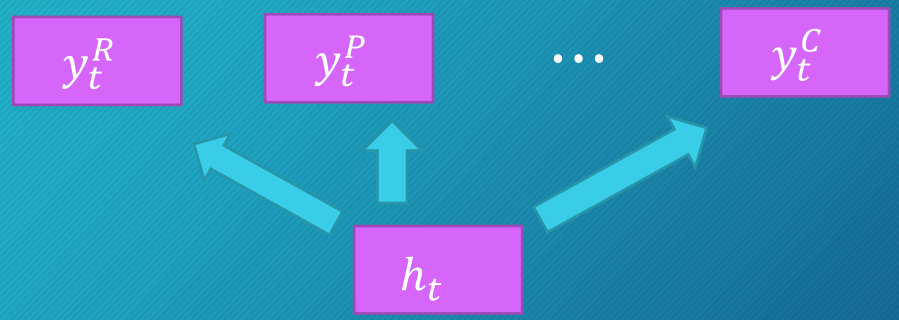
$$h_i = g(W_i^T h_{i-1} + a_i)$$

$$h_1 = g(W_1^T x + a_1)$$

- Parametric function
- Trained to reproduce input as output
- Complexity is constrained to prevent learning identity function
- Anomaly is detected when a poor reconstruction of an input is made by the model

Predicting Structured Events

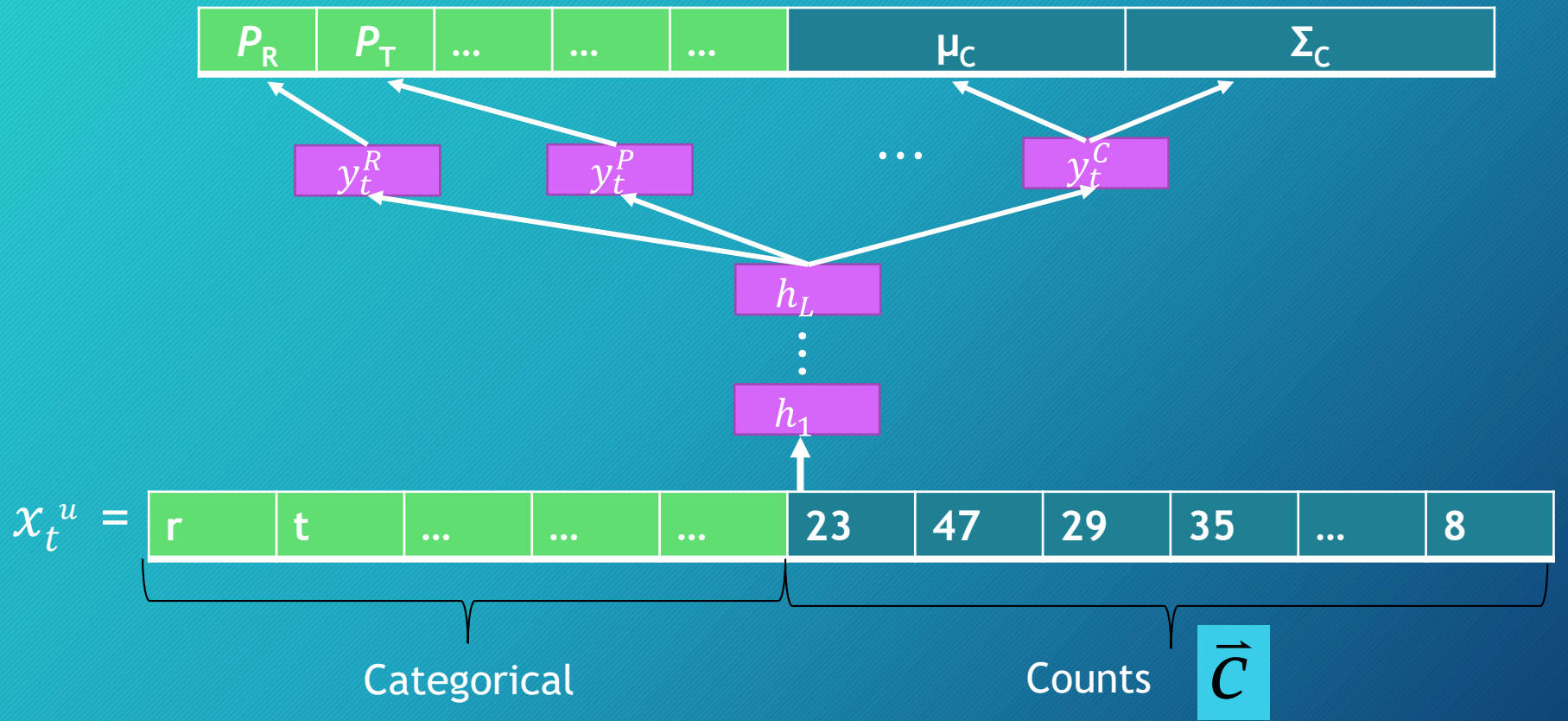
- First, we decompose the joint probability (assume independence)
  - $P(R, P, T, S, C|h_t) \approx P(Role|h_t)P(Project|h_t)P(Team|h_t) \cdots P(Counts|h_t)$
  - The output y vectors are either a distribution (for categorical input) or the parameters of a multivariate normal distribution (for vector of continuous input features).



Training and Anomaly Detecition

$$-\log(P_R(r)) + -\log(P_T(t)) + \ldots + -\log(P_C(\vec{c}))$$

Multivariate loss
(Anomaly score)



Outline

- Data processing and feature extraction
- Neural networks and our model
- **Experiments and results**
- Takeaways

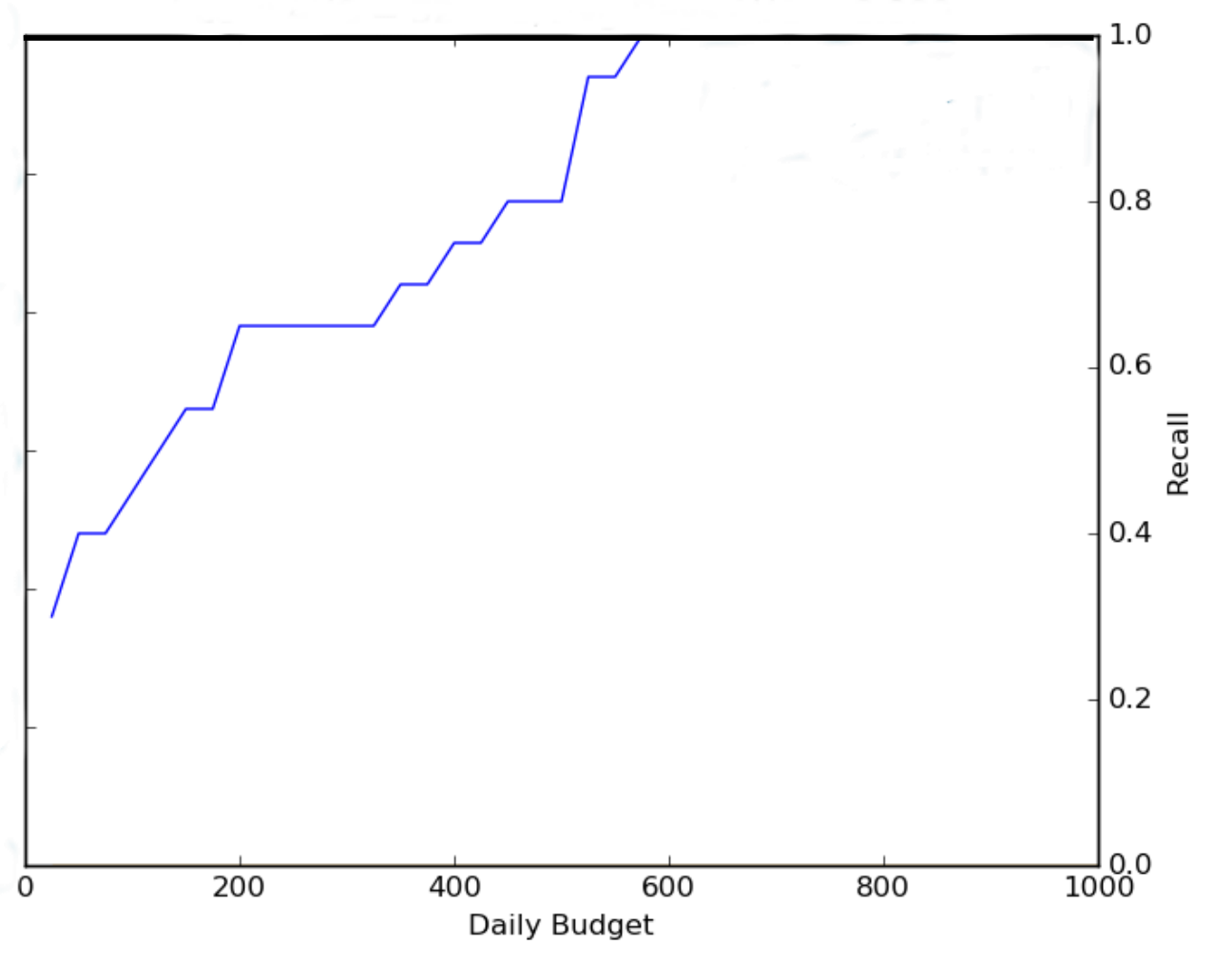
Experiment Setup

- Split data set into train/dev and test
- Aggregate Features: 408 count features, 6 categorical features
- Test model configurations on range of hyper parameters
- Test best model configurations against standard anomaly detection techniques

	Development	**Test**
Date Range	*Days 1 - 418*	*Days 419 - 516*
# Device Events	1,285,341	266,487
# Email Events	9,068,429	1,926,528
# File Events	1,671,698	343,185
# HTTP Events	96,516,038	20,509,178
# Logon Events	2,916,161	614,124
Total Events	**111,457,667**	**23,659,502**
Threat Events	192	236
Threat User-Days	27	20

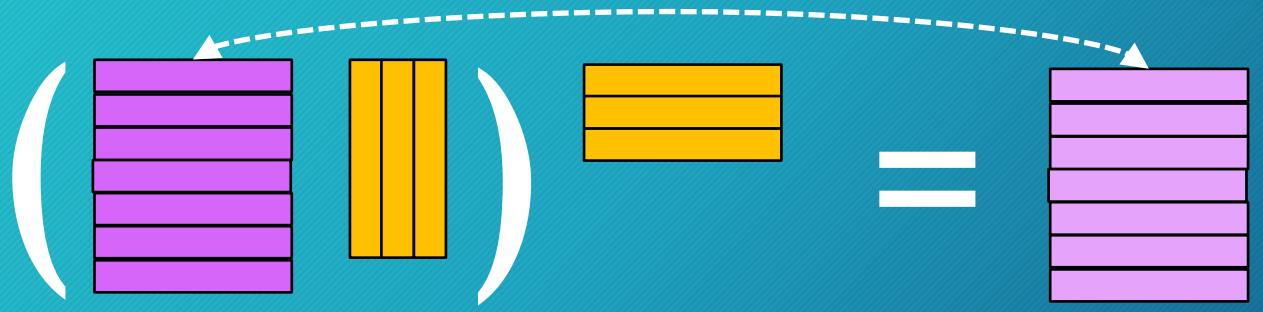
Evaluation Criteria

- CR-*k*: Sum of recalls for all budgets up to and including *k*

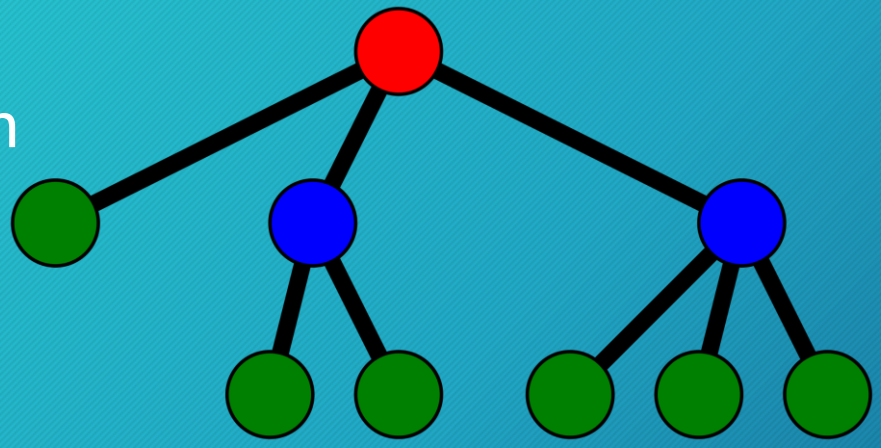


Baseline Models

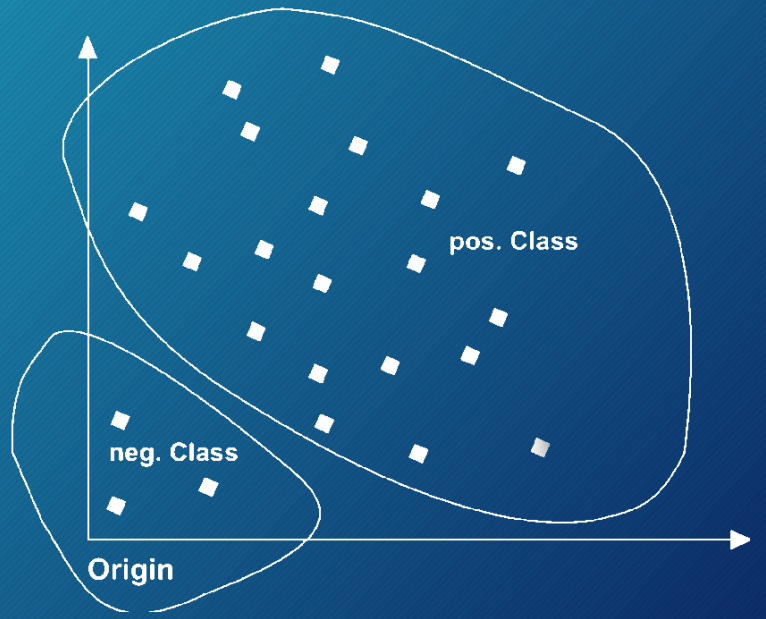
PCA Reconstruction



Isolation Forest

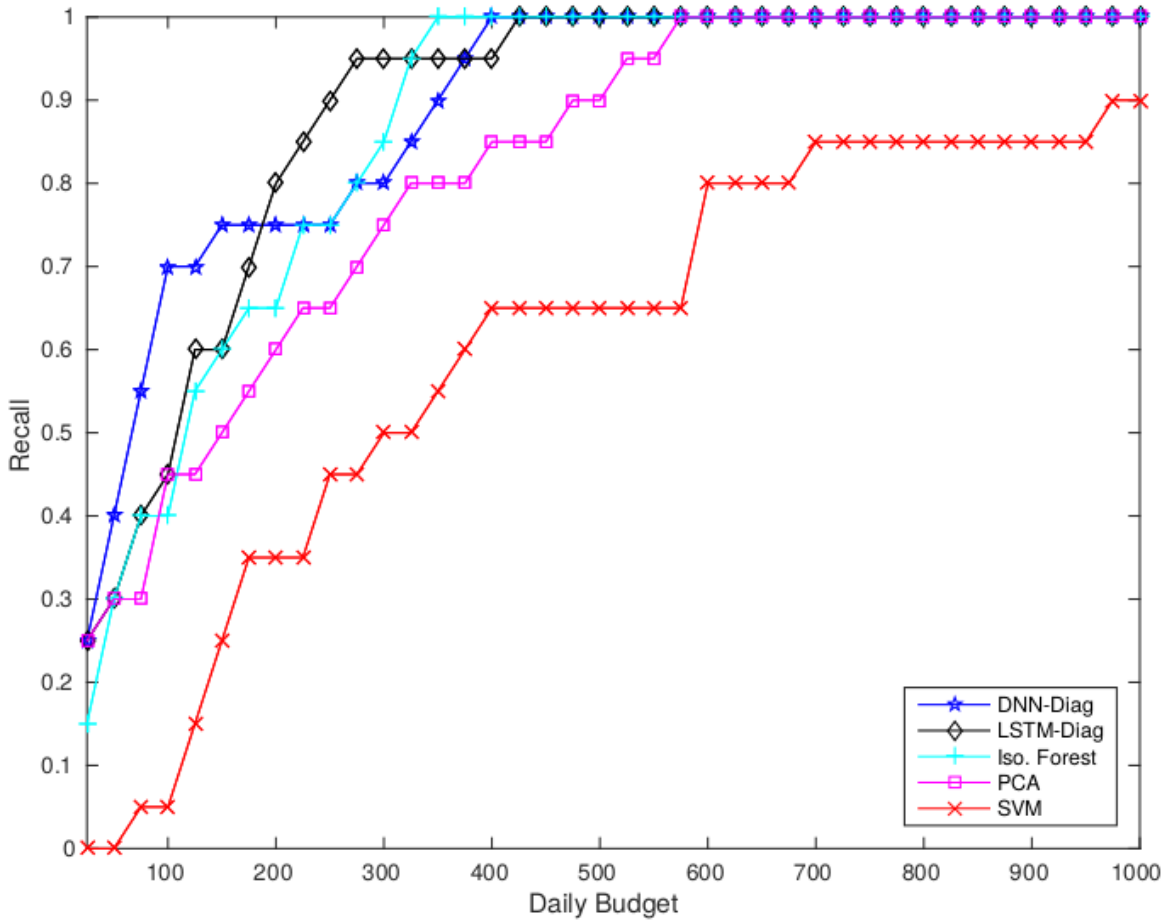


One-Class Support Vector Machine

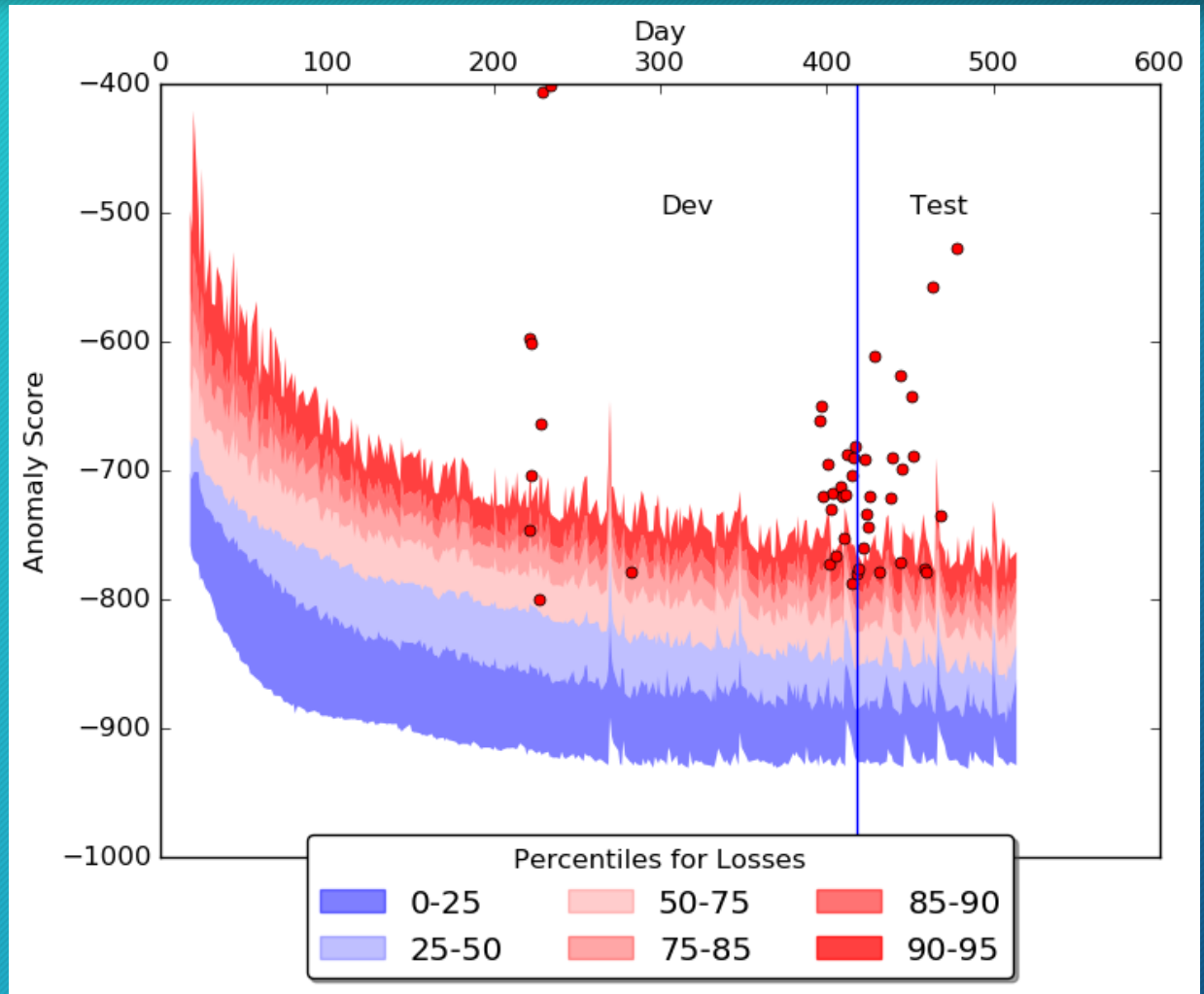


Dnn vs Rnn vs Baselines

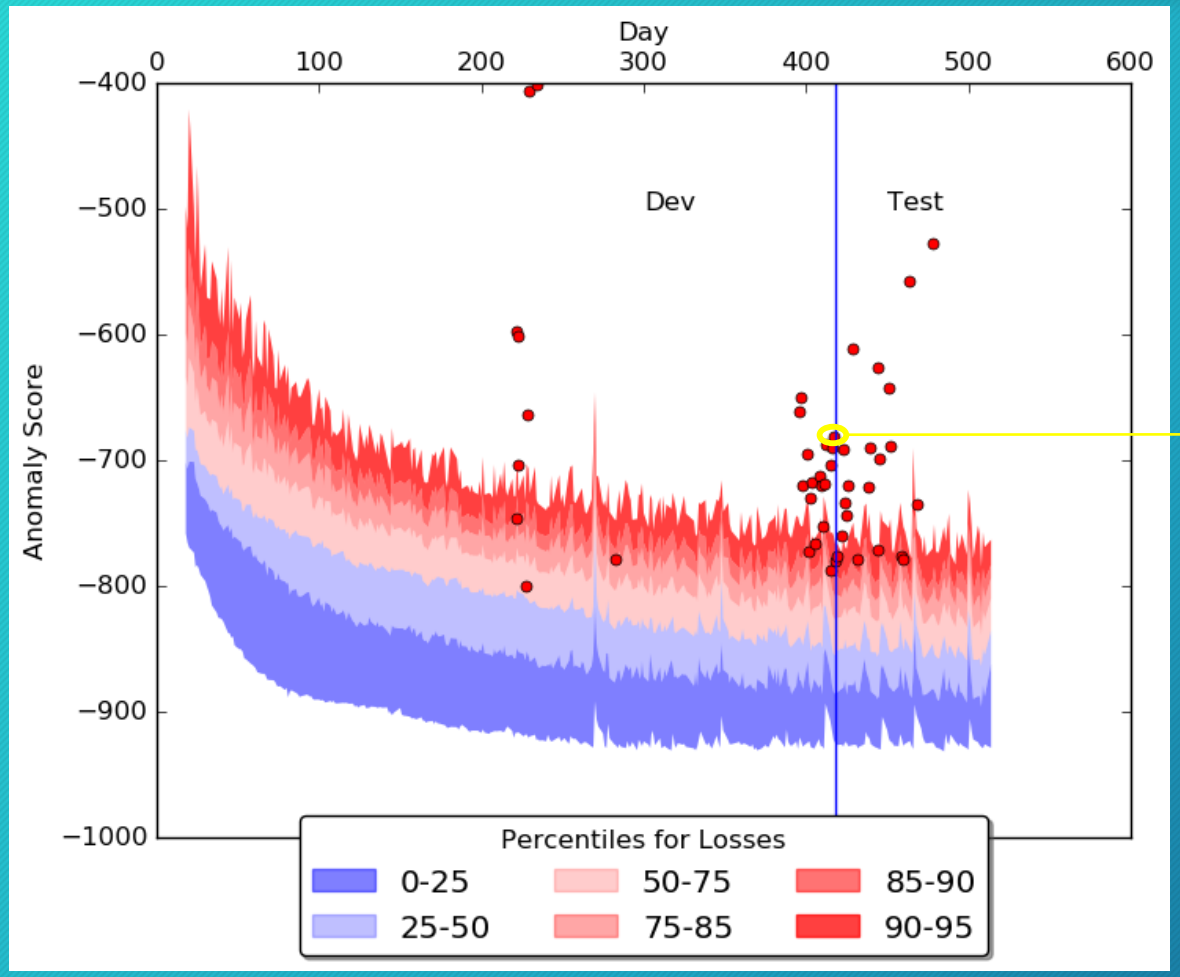
Model	CR-400	CR-1000
Isolation Forest	10.8	34.8
SVM	5.3	24.2
PCA	9.4	32.8
DNN-Ident	9.8	32.4
DNN-Diag	11.7	35.7
LSTM-Ident	10.8	33.0
LSTM-Diag	11.6	35.6



Best results: DNN-diag



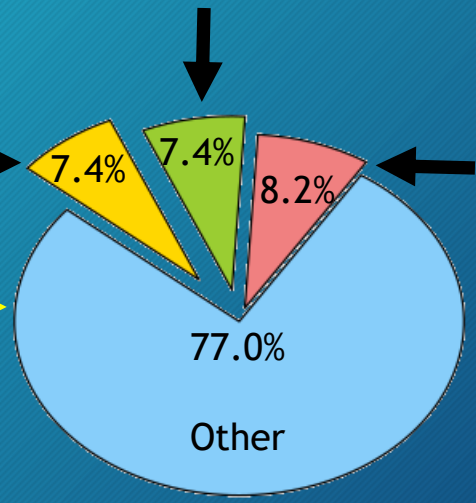
Interpretable Assessments



Write to a globally uncommon file locally 12pm-6pm

Copy to a globally Uncommon file from Removable media 6am-12pm

Write globally Uncommon file locally 6pm-12am



Day 418
True Positive

Decomposed negative log probabilities provide insight into anomaly score

Outline

- Data processing and feature extraction
- Neural networks and our model
- Experiments and results
- **Takeaways**

Takeaways

- Online unsupervised deep learning architecture
- Interpretable assessments
- System meets constraints of the online scenario
  - Assessments in real time
  - Bounded memory requirements
- System outperforms standard anomaly detection techniques
- Approach is applicable to a more general class of problems

Thank you

- Questions?