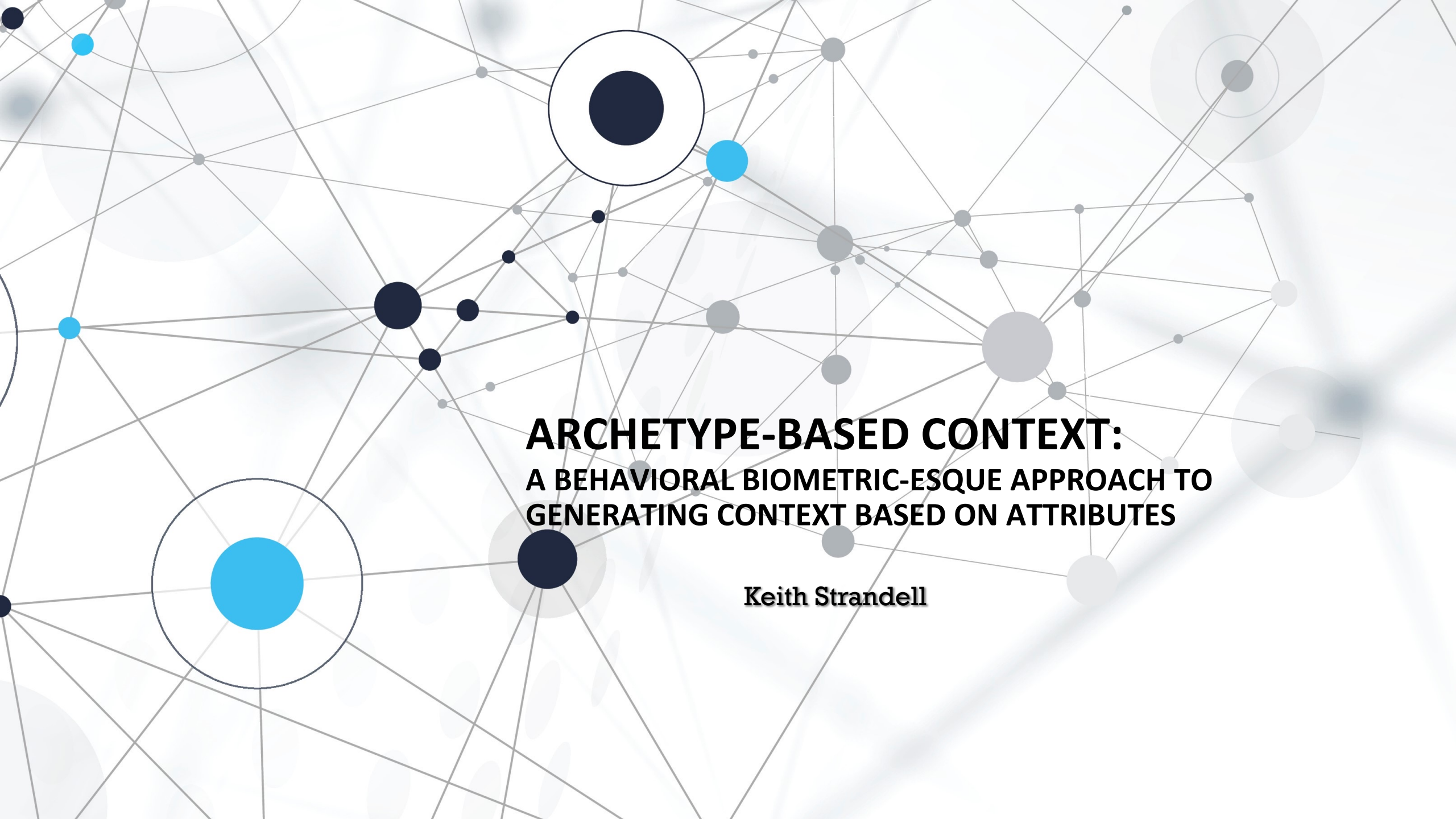


ARCHETYPE-BASED CONTEXT:
A BEHAVIORAL BIOMETRIC-ESQUE APPROACH TO GENERATING CONTEXT BASED ON ATTRIBUTES

Keith Strandell

DISCLAIMER

- The views expressed in the presentation are those of the author and do not necessarily represent those of the Air Force, the DoD, or the US Government.

As we move toward Attribute-Based Access Control and Zero Trust, the contextual information used to support authentication processes becomes increasingly useful in defining a robust trust algorithm. A Machine Learning based classifier, trained on audit logs paired with key user attributes can be used to provide a confidence level that a claimant is indeed the purported entity by asserting an expected archetype for comparison against the claimant's attributes. Consider the differences in how maintainers and IT specialists interact with IT systems or the commonalities seen due to operating location or organization. The brief will discuss archetype-based context and potential ML algorithms for use as classifiers.

POP QUIZ

Which one is a security risk?

POP QUIZ

Which one is a security risk?

1. Weather Memes. (n.d.). Pintrest. https://www.pinterest.com/pin/783204809904526 32/

AGENDA

- Who Am I?
  - Problem I'm researching
  - Why this problem space?
- Context
- Behavioral Biometrics
- Archetype-Based Context
  - Using Attributes to Generate Context
  - Notional Architecture
  - Context Augmentation Block - Models

WHO AM I?

- **Keith Strandell**
- Started in USAF as a Comm and Info Officer (back when 33SX still existed)
- Spent about 3.5 years in industry on programs supporting multiple countries
- Approx 11 years as civilian in AFLCMC
- Last 4 civilian roles:
  - (Foreign Military Sales) Materiel Leader for RAMP
  - (Information Technology) Materiel Leader for DAF365 / AF Portal / EMNS
  - (Foreign Military Sales) Program Manager – 12 cases (CENTCOM, PACOM and EUCOM)
  - (Information Technology) Section Chief – AF DCGS NETCOMMS

Currently pursuing a PhD in Computer Science at Mississippi State University

- Researching ways to enhance a Federated Zero Trust Architecture for enduring mission partner environments

PROBLEM SPACE

- Federated Zero Trust Environment dedicated to Collaboration and Data Sharing with Mission Partners
  - Executive Order 14028: Improving the Nation's Cybersecurity[1] directs transition to Zero Trust
  - Extension of NATO's Federated Mission Networking (FMN) concept
    - FMN built on Afghanistan Mission Network lessons learned – demonstrated benefits of multinational networking environment[2]
  - Need enduring network able to support rapid integration of new mission partners
    - Haiti Earthquake in 2010 highlighted issues w/ lack of data sharing capabilities[3]
    - Needs to be responsive to mission sets such as Humanitarian Assistance and Disaster Response (i.e., rapid integration)

1. Joseph Biden, "Exec. Order No. 14028," 3 C.F.R. 86 (May 2021).
2. "Federated Mission Networking." NATO Allied Command Transformation. https://www.act.nato.int/ (20 Jul 2023).
3. Gary Cecchine et al., The U.S. Response to the 2010 Haiti Earthquake (RAND Corporation, 2013).

WHY THIS PROBLEM SPACE?

Sharing data among mission partners is hard!	Changing international landscapes Information sensitivity (CUI, REL TO, …) Varying technology baselines → weakest link
Zero Trust moves the needle (in theory)	Smaller Inherent Trust Zones Continuous authentication Compartmentalized (better data segregation)
However	Federated models introduce risk (adding trust into Zero trust) Not all networks are created equal → weakest link Federation challenges with sharing context

CONTEXT

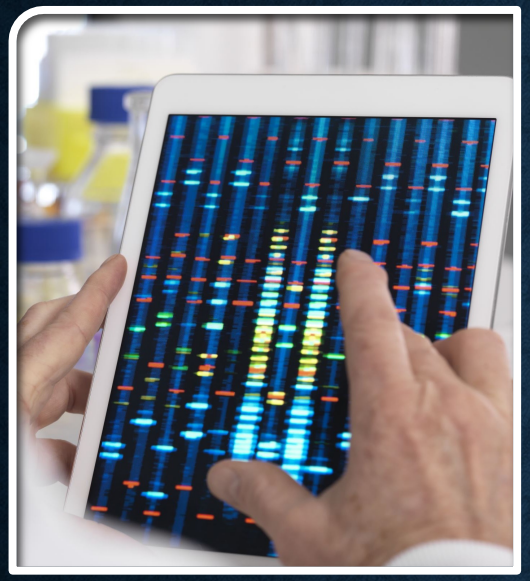
For me context is key – from that comes the understanding of everything

- Kenneth Noland

CONTEXT

- "The interrelated conditions in which something exists or occurs"[1]
  - For Trust Algorithms, it's the story around an access request
- Contextual Trust Algorithm – Takes recent history into consideration[2]
  - More data to support access decision
  - Paired w/Continuous Authentication – has potential to identify active compromise
- Access request without context
  - Bob requests access to F35 schedule data
- Access request with context
  - Bob, a program manager in the F35 program office requested access to F35 schedule data from an enterprise device at an IP address in Washington DC as he does every Monday at 1000
  - Bob, a contractor in a services squadron requested access to F35 schedule data from a non-enterprise device at an IP address in Argentina and over the past 5 hours has downloaded a dozen related engineering documents

1. "Context." Merriam-Webster.com. 2023. https://www.merriam-webster.com (8 July 2023).
2. Scott Rose et al., "NIST Special Publication 800-207," August 2020,



BEHAVIORAL BIOMETRICS

- How does a user interact with the system
  - Keyboard Dynamics
    - Any hunt and peck typers here?
  - Mouse Dynamics
  - Touch Dynamics
  - Multimodal
  - Browsing / App Usage

ARCHETYPE-BASED CONTEXT

ARCHETYPE-BASED CONTEXT

- Why look at archetypes?
  - Target system has potential for highly transient population
  - Limited / Insufficient data on individual users
  - Many user specific biometrics require client-side agent
  - Server-side data supports generalized assessments
- What can we expect to find?
  - Macro bounds on group behaviors
  - Micro shifts of individual within the larger bounds
- What are we trying to find out:
  - Does a maintainer interact differently with the system than a PM? What about a mission partner's operator compared to the US counterpart?
  - **Simply put:** Are there common factors such as location, job function, shift, experience, … that impact a user's behaviors
    - Frequency of use
    - Applications used
    - Devices used

USING ATTRIBUTES TO GENERATE CONTEXT

- Requirements
  - Standardized, stable attribute model
  - Ability to query attributes for a given request
  - Attributes tied to logs (likely using one-hot encoding)
    - User attributes change over time – changes in attributes can change context
    - Avoid arbitrary ordinality

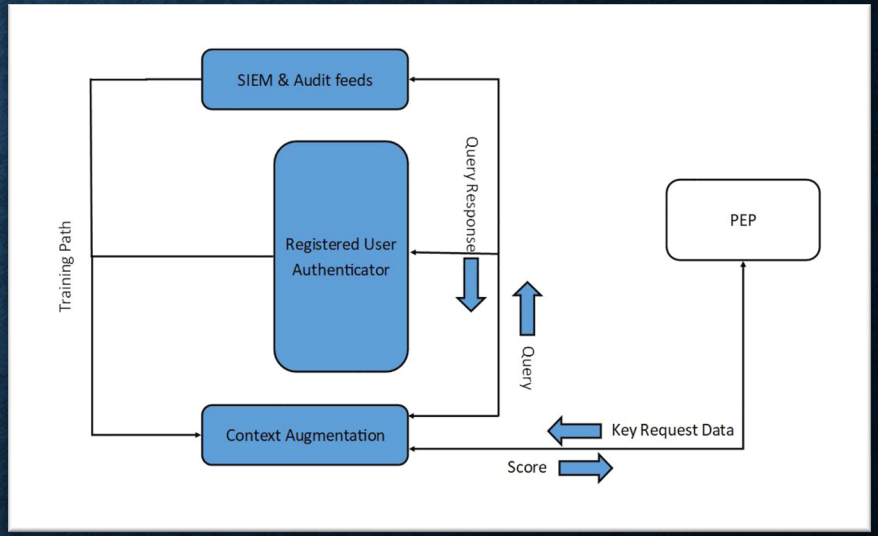
Job Function					Time Zone			Shift Work		
Program Management	Engineering	Manufacturing	Contracting	Finance	EST	PST	MST	First	Swing	Graveyard
1	0	0	0	0	0	0	1	1	0	0

- can be stored as 1036

  - Access to SIEM and audit data
    - Provides the historical information used to generate context whether individual or archetype based
    - Source for training data

NOTIONAL ARCHITECTURE

- ID Management System (shown as Registered User Auth)
  - Source of truth on user identity
  - **Standardizes attributes** – key for archetype-based context
- SIEM & Audit Feeds
  - Provides the history to generate context
  - Store system logs along with user data
  - May need method of tagging data (e.g., confidence the event was valid)
    - Ability to provide tag and provide data on malicious or less than ideal attempts
- Context Augmentation
  - Houses instances of the assessment model(s)
    - Separate operational and training models
  - Receives data from PEP and queries appropriate feeds for data
  - Uses context to generate score(s) to be provided to Trust Algorithm



CONTEXT AUGMENTATION BLOCK - MODELS

- Statistical assessment
  - Can be automated to a degree
  - Enumerated rule set – new rule for each discovery
  - Potential for human bias
- Machine Learning
  - May identify subtle behaviors
  - Can handle large data sets
  - Can support continuous training model
  - Lack of visibility – black box
  - Potential for human bias in selecting / formulating data sets

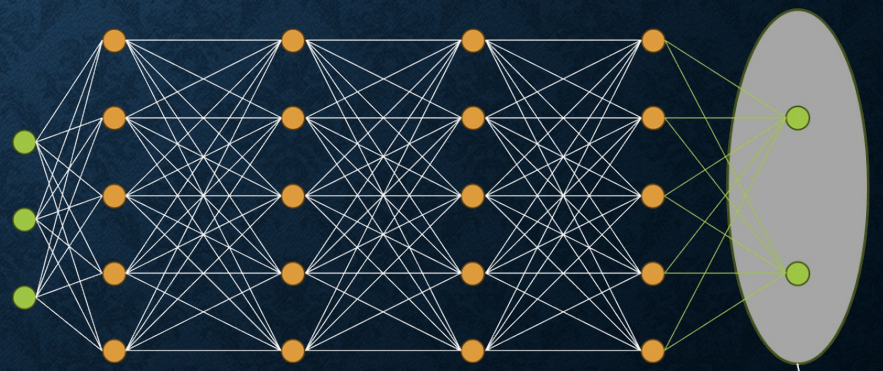
CONTEXT AUGMENTATION BLOCK – MODELS CONT'D

- Specific Machine Learning models to consider
  - Deep Artificial Neural Network
    - Can support multi-modal (macro / micro assessments)
    - Useful in generalized, complex problems
    - Can serve as classifier
  - Adaptive Neuro Fuzzy Inference System
    - Multilayered neural network that optimizes a fuzzy inference system
    - Demonstrated ability to model non-linear behavior
    - Study implemented ANFIS for continuous authentication by profiling a user based on several features, including app history[1]
  - Long Short-Term Memory Network
    - Form of Recurrent Neural Network
    - Uses current and previous state in assessment
    - May prove beneficial in continuous authentication

1. Yao et al., "Continuous implicit authentication for mobile devices based on adaptive neuro-fuzzy inference system."

CLASSIFIER W/DANN

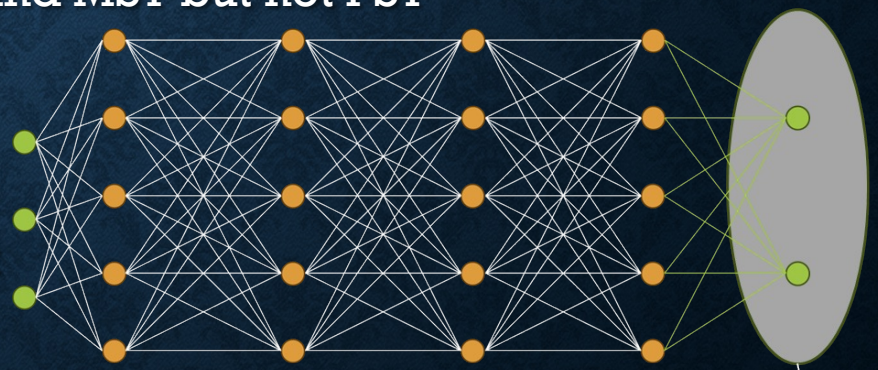
- Training
  - Utilizes the preprocessed data of a user's request and provides a guess
  - The guess is compared with the answer (in this case the one hot encoding of the user's attributes)
  - Error is assessed and backpropagation is used to make minor adjustments
- Testing
  - A similar data set is used to gage the accuracy
  - No backpropagation – model is locked



Job Function					Time Zone			Shift Work		
Program Management	Engineering	Manufacturing	Contracting	Finance	EST	PST	MST	First	Swing	Graveyard
1	0	0	0	0	0	0	1	1	0	0

CLASSIFIER W/DANN

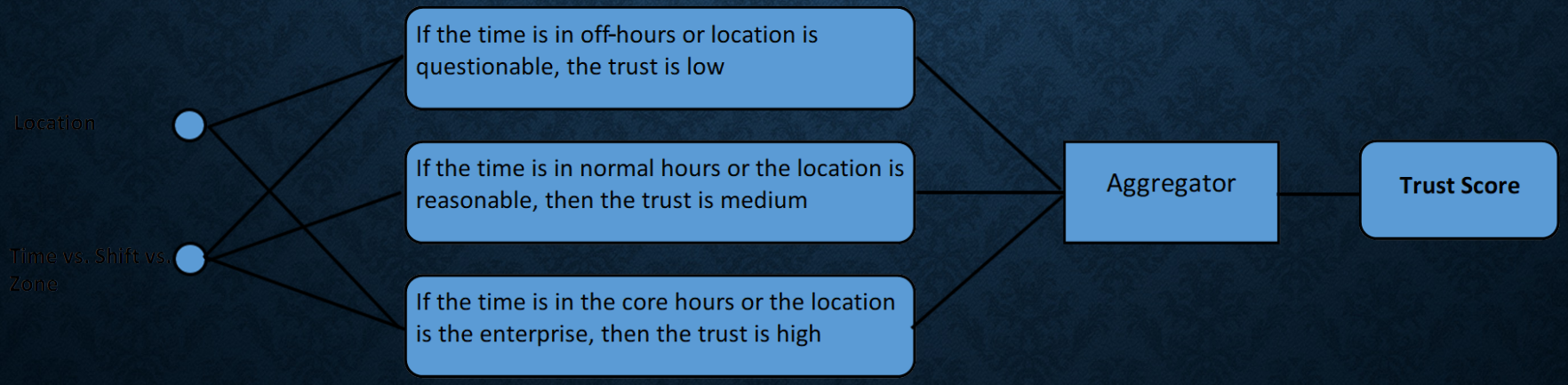
- Classifier will not provide a binary answer for each attribute
- Instead, it will provide a value x, such that $0 \leq x \leq 1$
- Compare results with the user's actual attributes (ignore the zeros)
  - We are interested in the *reasonableness* of a user's attributes
  - Ex. A timestamp of 1300Z may be high for both EST and MST but not PST
- Challenges
  - How to define reasonableness
  - The spread of values will be different for every input



Job Function					Time Zone			Shift Work		
Program Management	Engineering	Manufacturing	Contracting	Finance	EST	PST	MST	First	Swing	Graveyard
1	0	0	0	0	0	0	1	1	0	0

FUZZY INFERENCE SYSTEM

- Takes inputs and uses ambiguous rules to make a determination
- Rules are typically defined by a SME; however…
  - There are models that can learn the rules from the data
    - E.g., Adaptive Neuro Fuzzy Inference System



CLOSING

- Networks can produce a significant amount of data that can be used to support authentication decisions
- Strong attribute models, coupled with request data can be assessed to generate additional context
  - e.g., a reasonableness score by training a classifier and defining a means of comparison against a user's attributes
- Various machine learning models can be utilized to identify patterns in the data to support making these determinations

QUESTIONS