

# LINEBACKER: LINE-speed Bio-inspired Analysis and Characterization for Event Recognition



## **Cyber Security Division 2012 Principal Investigators' Meeting**

10/11/2012

**Christopher Oehmen  
Sr. Research Scientist  
Pacific Northwest National Laboratory  
Christopher.Oehmen@pnnl.gov  
509-375-2038**

# Introduction

- TTA 13
- Key staff and collaborators
  - Chris Oehmen (PNNL), PI, Task lead for sequence analysis algorithm design and implementation
  - Bill Pike (PNNL), Task lead for visual analytics
  - Bora Akyol (PNNL), Subject matter expert for network traffic analysis
  - Doug Pearson (REN-ISAC), Director of REN-ISAC and integration lead

# Introduction to LINEBACKER

Network traffic is complex, diverse

- No obvious global “normal”

Network traffic is hierarchical

- Packets may be ok, but in sequence might be recognizably bad

Network traffic is dynamic

- Normal, malicious tomorrow will drift from how they look now

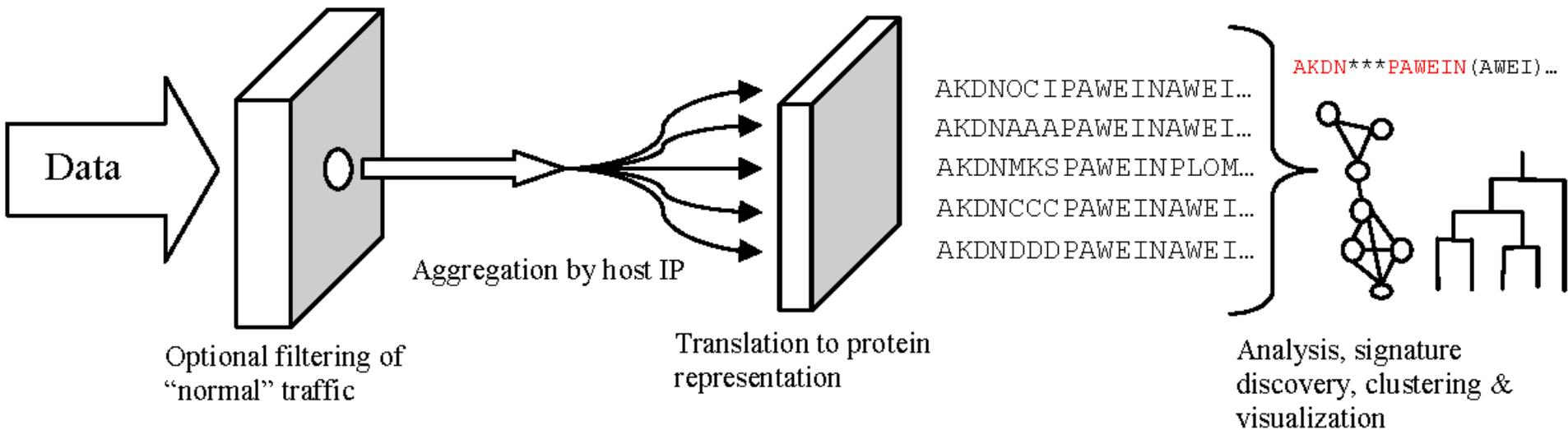
Data is mainly diagnostic, and was originally designed for troubleshooting, not security

- 1) Given all the activity on my network now, what is the space of behaviors that are happening?
- 2) Given a particular behavior, what does it look most like?

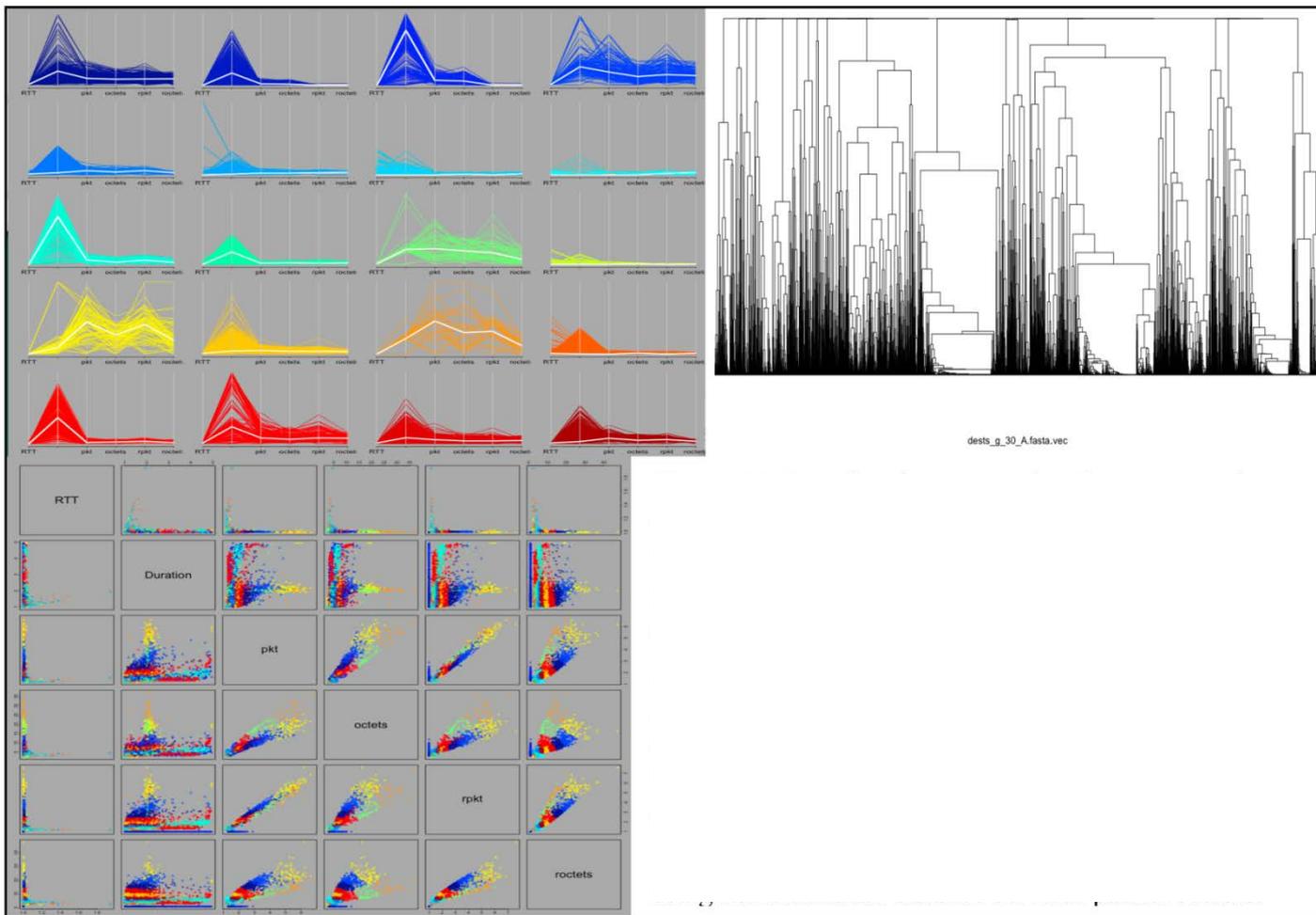
# LINEBACKER solution

- LINEBACKER concept
  - use a transformed representation of packet *sequences*
  - associate sequences each other to suggest *families*
  - build *models* of the family from highly conserved attributes of family members
  - use models as the basis for new *signatures/sensors*
- Can be constructed using normally available data
- Patterns emerge from quantitative analysis, suggest new signatures to look for
- Can account for drift in the underlying phenomena using biological principles

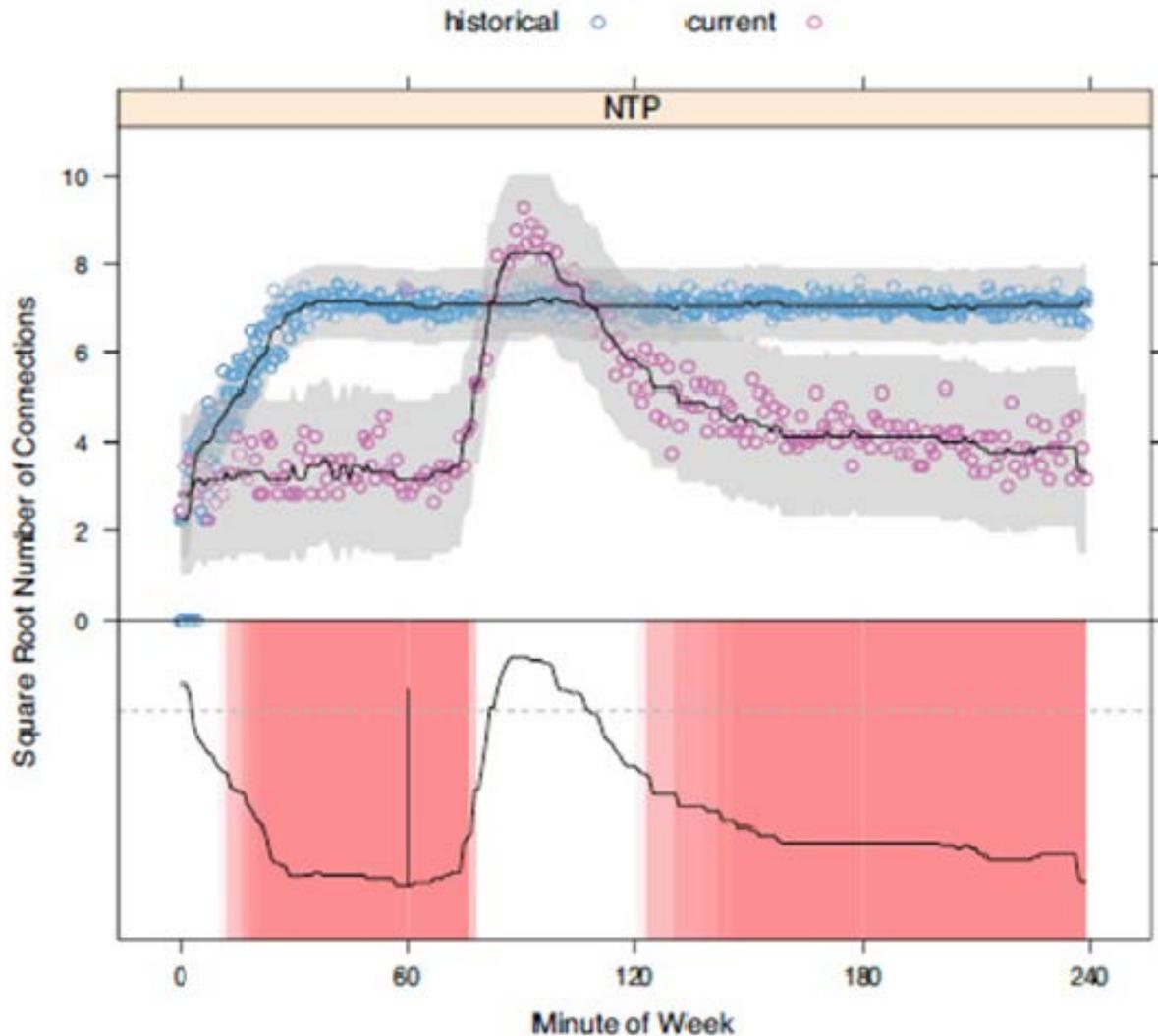
# Overview



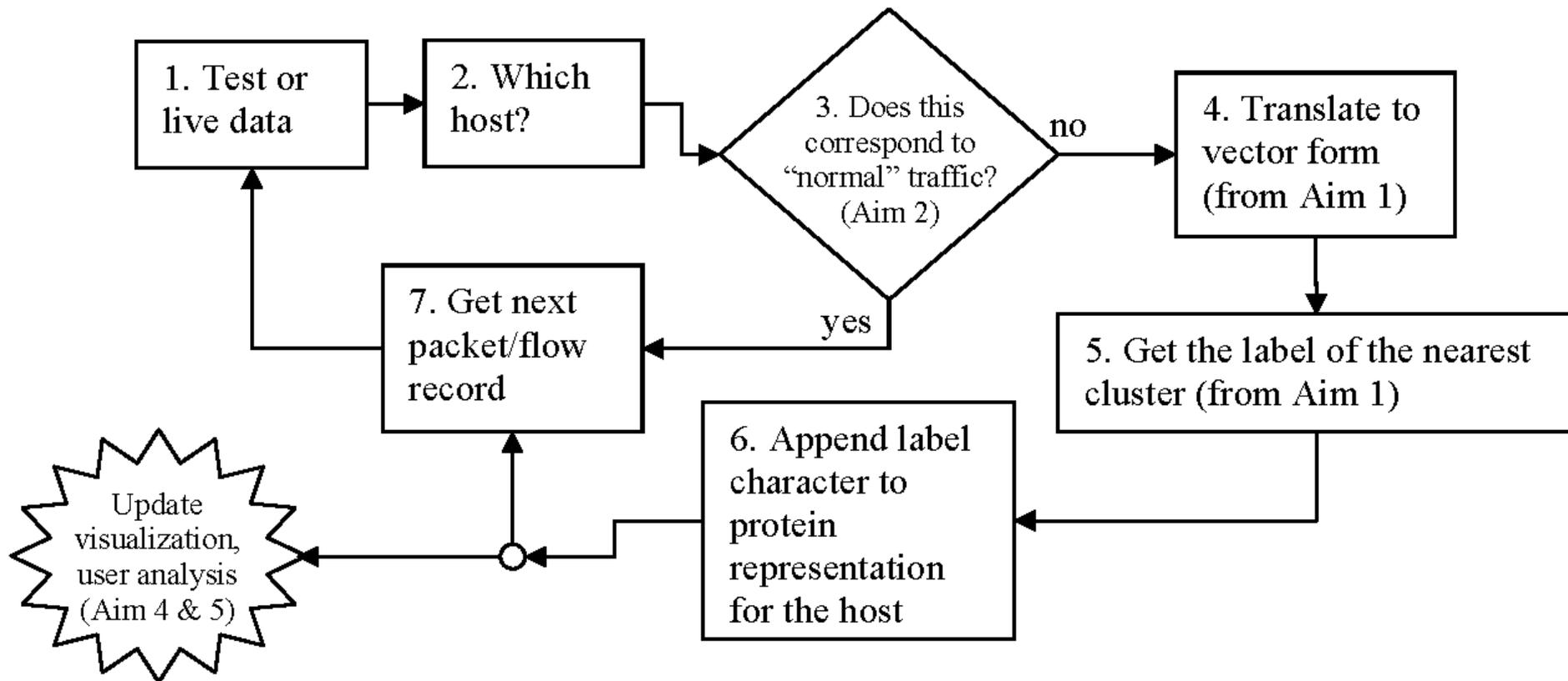
# Transform traffic to sequences



# Associate sequences with behavior



# LINEBACKER process

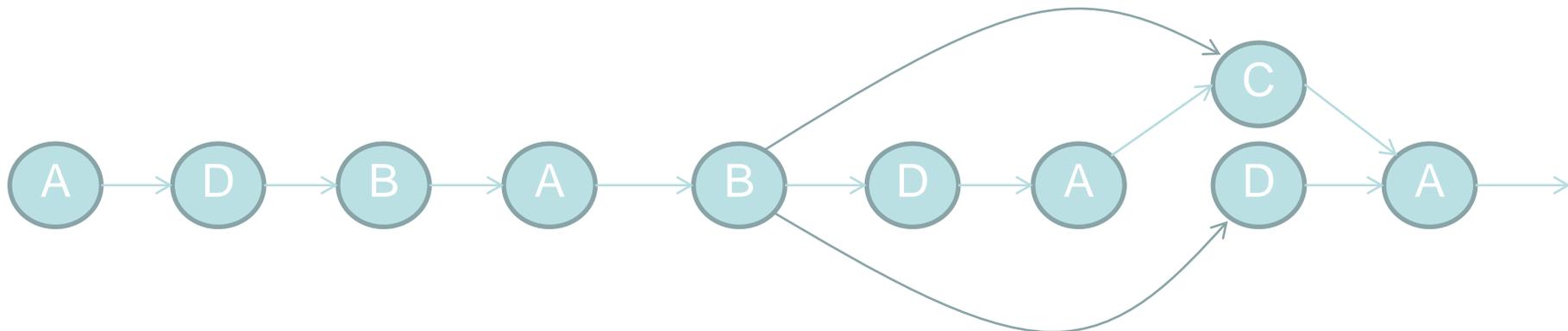


Query sequence : ABCDDBDBBDDDBAAABABDBACCC--BCBDBABBBDDDBBBDDD  
**Consensus region** DBAAABABD ACCC BCXDBA  
 Target sequence: BCCCCBAAAAADBAAABABD-ACCCDABCADBAAAACCCAAACCC

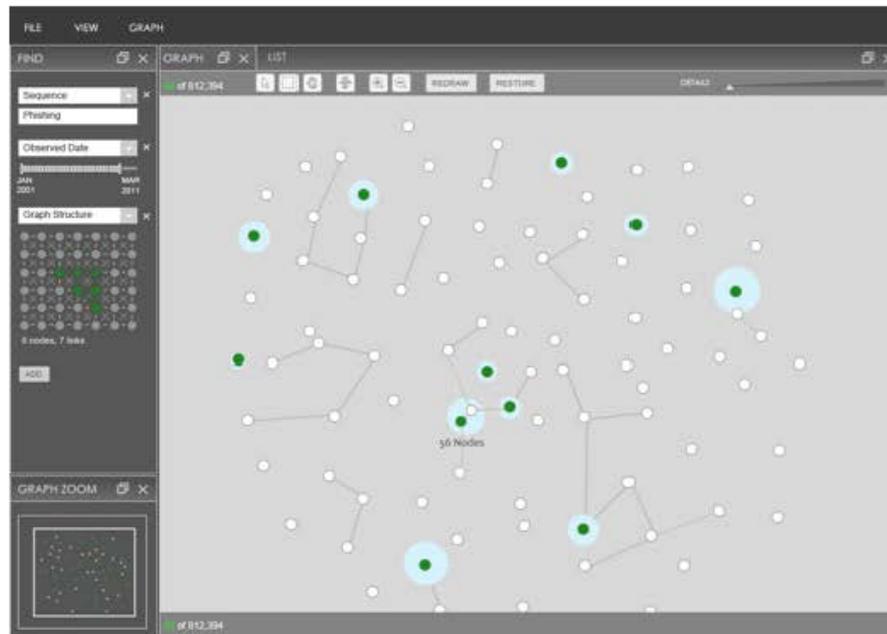
# Generating Signatures

- Sequence 1: **ADBAB**DAC**ACBADCCBACBDBCDDDBCDB**CBBCBCB...
- Sequence 2: **ADBAB**DAC**ACBADCCBACBDBCDDDBCDB**CBBCBCB...
- Sequence 3: **ADBAB**--C**ACBADCCBACBDBCDDDBCDB**CBBCBCB...
- Sequence 4: **ADBAB**--C**ACBADCCBACBDBCDDDBCDB**CBBCBCB...
- Sequence 5: **ADBAB**--D**ACBADCCBACBDBCDDDBCDB**ADADADA...
- Sequence 6: **ADBAB**--D**ACBADCCBACBDBCDDDBCDB**ADADADA...

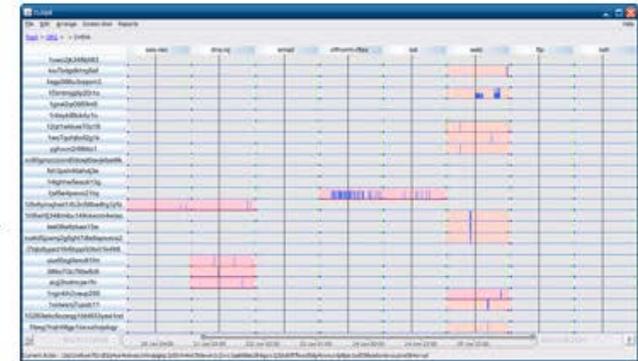
Consensus : **ADBAB      ACBADCCBACBDBCDDDBCDB**



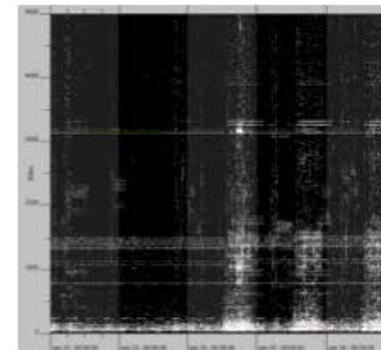
# Visualization



Sequence family visualization



Visualization of sequence temporal patterns



Visualization of raw traffic comprising a sequence

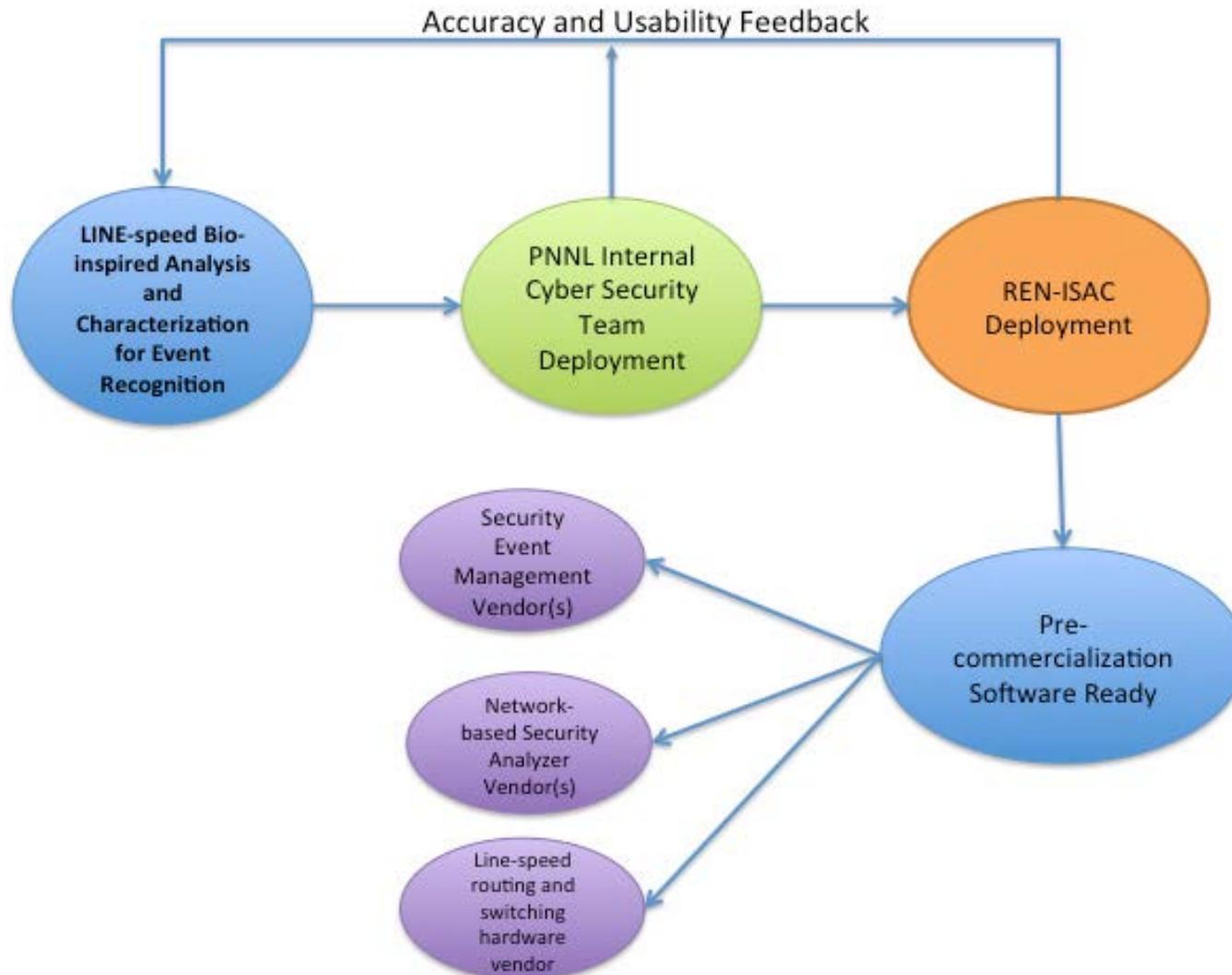
# Milestones

Aim	Description
1	<ul style="list-style-type: none"><li>• Identification of attributes that lead to high-quality clusters of behavior primitives</li><li>• Cluster definitions and labels for letterization</li></ul>
2	<ul style="list-style-type: none"><li>• Development of baseline behavior models</li></ul>
3	<ul style="list-style-type: none"><li>• Construction of software framework for rapid translation of network traffic</li></ul>
4	<ul style="list-style-type: none"><li>• Construction of a family tree for network traffic</li><li>• Identification of sequence motifs for network traffic from family tree</li></ul>
5	<ul style="list-style-type: none"><li>• Construction and deployment of visual/analysis framework</li></ul>
Validation and testing	<ul style="list-style-type: none"><li>• Deployment of framework and training at REN-ISAC deployment site</li><li>• Statistical performance and run-time efficiency evaluation of framework at REN-ISAC deployment site</li><li>• Design and execution of experiments to be run using DETER as a secondary platform</li></ul>

# Deliverables and timeline

Deliverable	Date	Description
Clique V1.0 to REN-ISAC	90 days	Provide initial baseline modeling capability for data managed by REN-ISAC, to provide basis for removing normal sequences from analysis
Sequence Detection V 1	411 days	Interactive system to classify traffic, remove baseline data, and identify cyber motifs
Sequence Detection V 2	548 days	Capabilities added to LINEBACKER product for whitelisting normal traffic, string matching, and similarity scores between cyber motifs
Visual Interface	579 days	User-facing GUI to explore cyber genes and families
Installation, Training, Document.	684 days	Training courses for REN-ISAC users completed; integration of sequence indicators into REN-ISAC SES completed and user/admin documentation delivered.
Reporting	Monthly, periodic	Monthly, annual, and final reports for project, attendance at PI meetings

# Technology Transfer



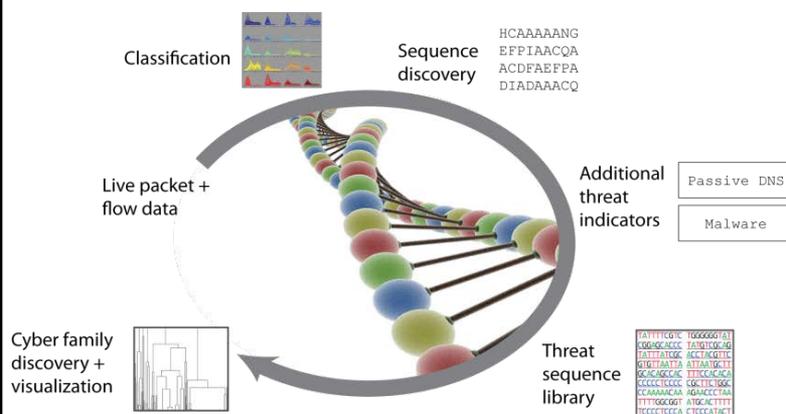
•Cyber Security BAA 11-02

•**LINEBACKER: LINE-speed Bio-inspired Analysis and Characterization for Event Recognition**

•Biosequence-based discovery of evolving threats

Pacific Northwest National Laboratory

TTA# 13



**Operational Capability:**

1. Ability to discover malicious network activity through sequence analysis across the U.S. research and engineering computing infrastructure
2. Construction of sequences from packet/flow data at rates exceeding 10 billion records per day
3. Support for submission and correlation of sequence patterns in the Research and Education Network Information Sharing and Analysis Center (REN-ISAC) Security Event System

**Proposed Technical Approach:**

1. Apply high-performance biosequence analysis that enables inexact string matching of streaming network traffic. Approach is robust to polymorphic threats and supports “family resemblance” attribution.
2. **Tasks:** Characterize baseline behavior, convert raw packets to bio-representation, construct family tree of cyber event types, create visual interface, deploy at REN-ISAC for the Global Research Network Operations Center
3. **Current status:** Builds upon existing MLSTONES (TRL 3) and CLIQUE (TRL 7) applications

**Schedule, Cost:**

Type II (2.5 yr)

**Deliverables:**

Operational product/tech transfer of biosequence-based threat detection for use in 300+ institutions collaborating via REN-ISAC

Ability to deliver capability to US-CERT as part of existing operational relationship

**Corporate Information:**

Pacific Northwest National Laboratory

Christopher Oehmen

PO Box 999, MSIN J4-33, Richland, WA 99352

(509) 375-2038; email: christopher.oeihen@pnl.gov