

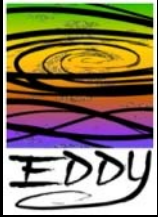
# The EDDY Initiative

Progress to Date

**Chas DiFatta (chas@cmu.edu)**

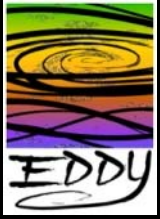
**Mark Poepping (poepping@cmu.edu)**

**Carnegie Mellon**



# Outline

- Motivation
  - The EDDY approach
  - Architecture and technical challenges
- EDDY enabled applications
  - Current and planned efforts
- Release 0.5
- Current Activities



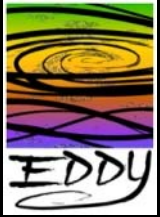
# Problem Example

You discover your bicycle has a flat tire...

- You fix it you move on

You discover another flat tire a week later...

- Valve problem?
- Nail in driveway
- Neighbor kid busting my chops?



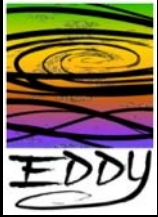
# Problem Example Cont.

You discover your Sendmail daemon crashed...

- You restart it and you move on

You discover it crashes a week later...

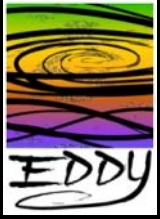
- Configuration problem?
- Performance or resource problem?
- New bug or integration problem with spam engines?
- Security vulnerability? Is it “really” sendmail running or a rogue daemon?



# Problem

## Banes of the Distributed System Diagnostician

- Limited **access** to slices of diagnostic data
- **Discovering** valuable information in a sea of data
- **Correlating** different diagnostic data types
- Providing evidence for **non-repudiation** of a diagnosis
- Finding **time** to create tools to transfer diagnostic knowledge to less skilled organizations and/or individuals (automation)



# Who are the Distributed System Diagnosticians?

Applications Support Personnel

Systems Administrators

Network Support Staff

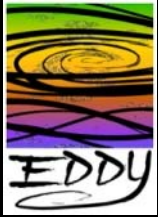
Security Response Practitioners

Managers of Computing Infrastructure in other  
Domains

Physical Infrastructure Managers/Engineers

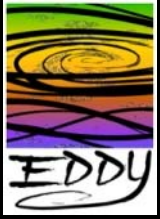
Help Desk

Ordinary Users

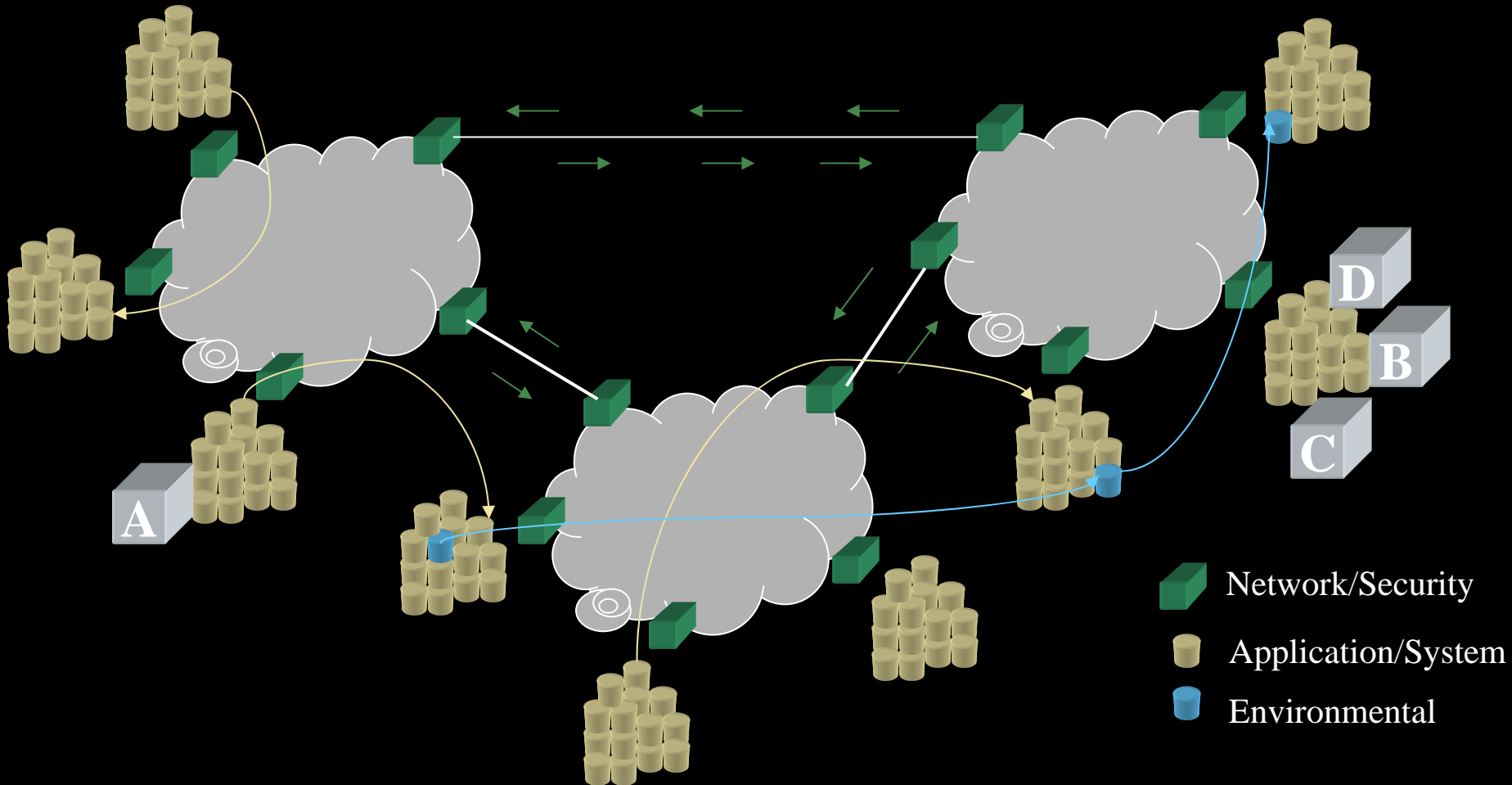


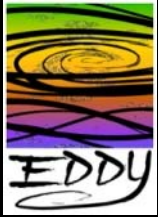
# State of Practice

- Each domain has its own tools and data collection methods, little cross-domain correlation
  - Example: Web error-log but a DNS configuration problem
- Most data represents only what has faulted
  - Example: web service or route down
- No end-to-end accountability of transactions. e.g. email, web, VoIP, intrusion
  - Example: Web access-log, no 2<sup>nd</sup> tier service, network, DB, etc.
- Scale is a serious issue with the proliferation of high density processors and embedded systems on high-speed networks



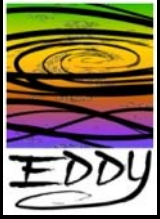
# Separate Event Domains





# Problem

- A Network diagnostician is asked to solve a problem with a physics application of researcher A
- It normally runs in 10 min prior to last Wednesday
- On Wednesday and afterward it took 60 min
- What is happening?



# Separate Event Domains

Network

Netflow (DNS)

NetFlow (Kerberos)

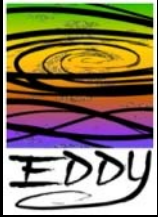
NetFlow (22:src=C, dst=D)

NetFlow (80:src=A, dst=B)

NetFlow (22:src=C, dst=D)

NetFlow (80:src=A, dst=B)

⋮



# Separate Event Domains

Application/System

Kerberos (uid=A)

Start Backups

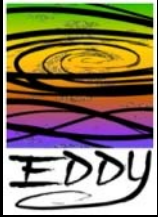
Start Physics App

Finish Backups

.

.

Finish Physics App



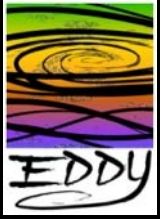
# Vision

Create an activity audit ledger that...

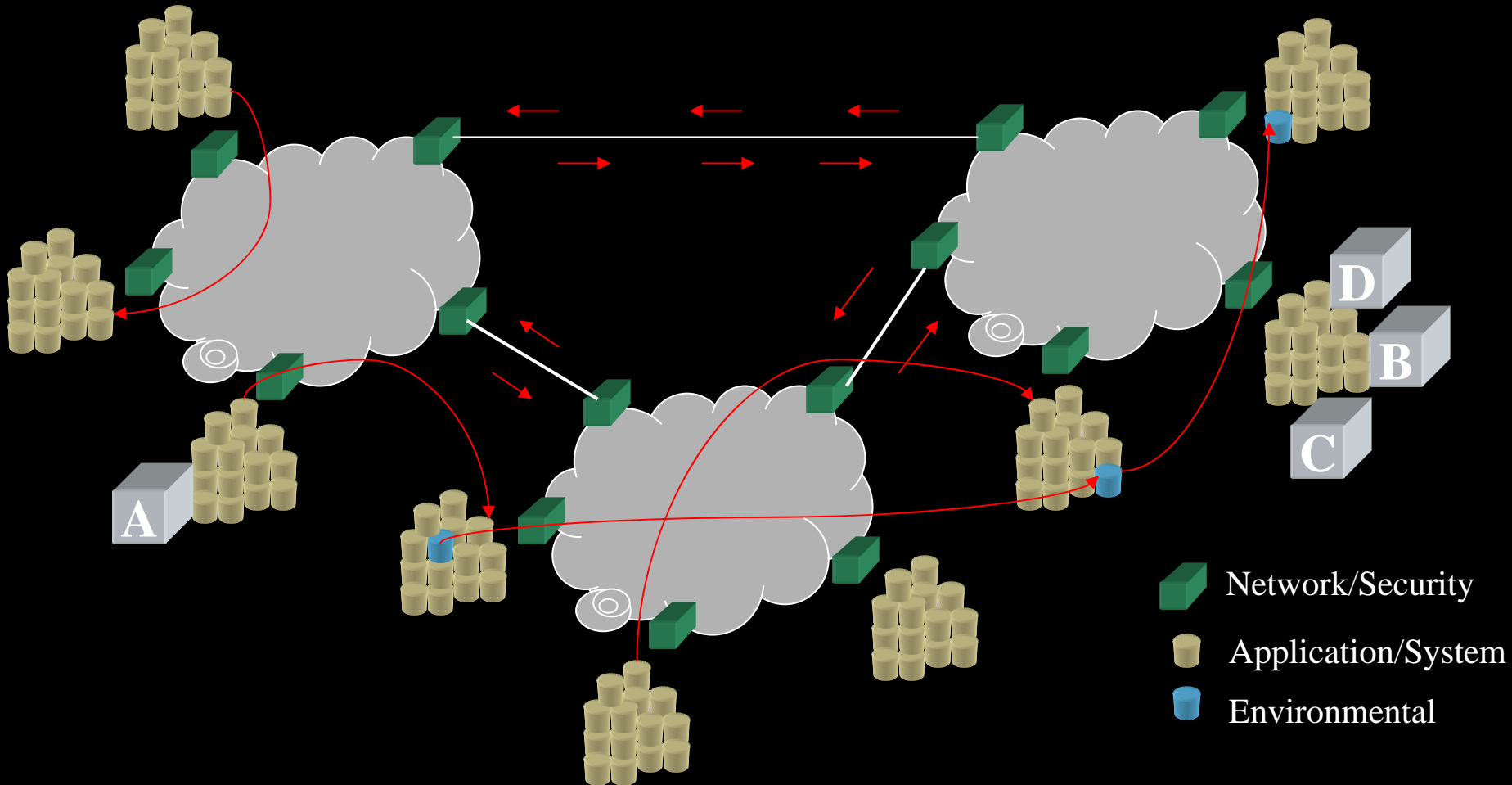
- Provides telemetry to study the behavior of layered and interdependent services, especially faults, anomalies, and in situ usage patterns
- Enables exploration of an Internet with auditable electronic communications and the influence on infrastructure, security, reliability, privacy and trust

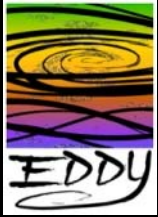
Provide a unified and robust orchestration infrastructure for audit ledger information to...

- Serve as a flexible substrate
- Intra- and inter-domain diagnostic, forensic, and performance analytics
  - Available for long-range research, but immediately useful for today's engineering problems

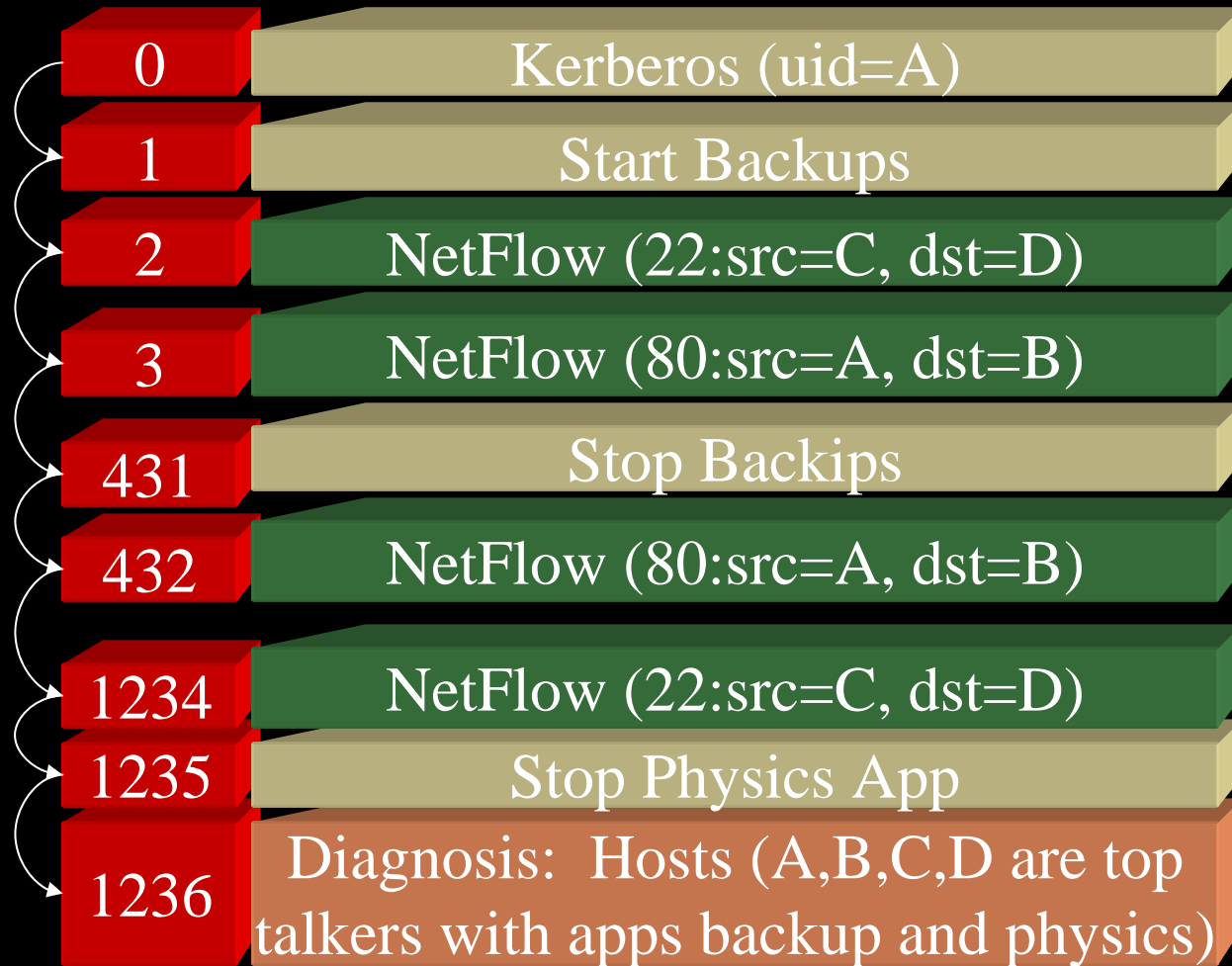


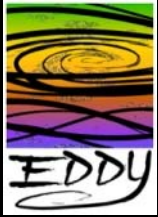
# Combined Event Domains





# Combined Event Domains

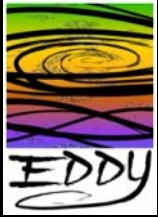




# Initial Direction

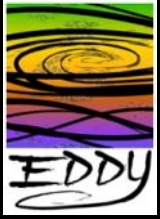
Enabling mechanism for investigating:

- Host to host interaction
  - Profiling behavior
  - Root cause forensics of faults
- Automated diagnostic practices, not just what has faulted but how the fault occurred
- Perceived anomalies verses actual faults
- High volume/density event clusters to embedded system events
- Taxonomic risk analysis of security anomalies
- Rapid tool development platform for diagnostic applications

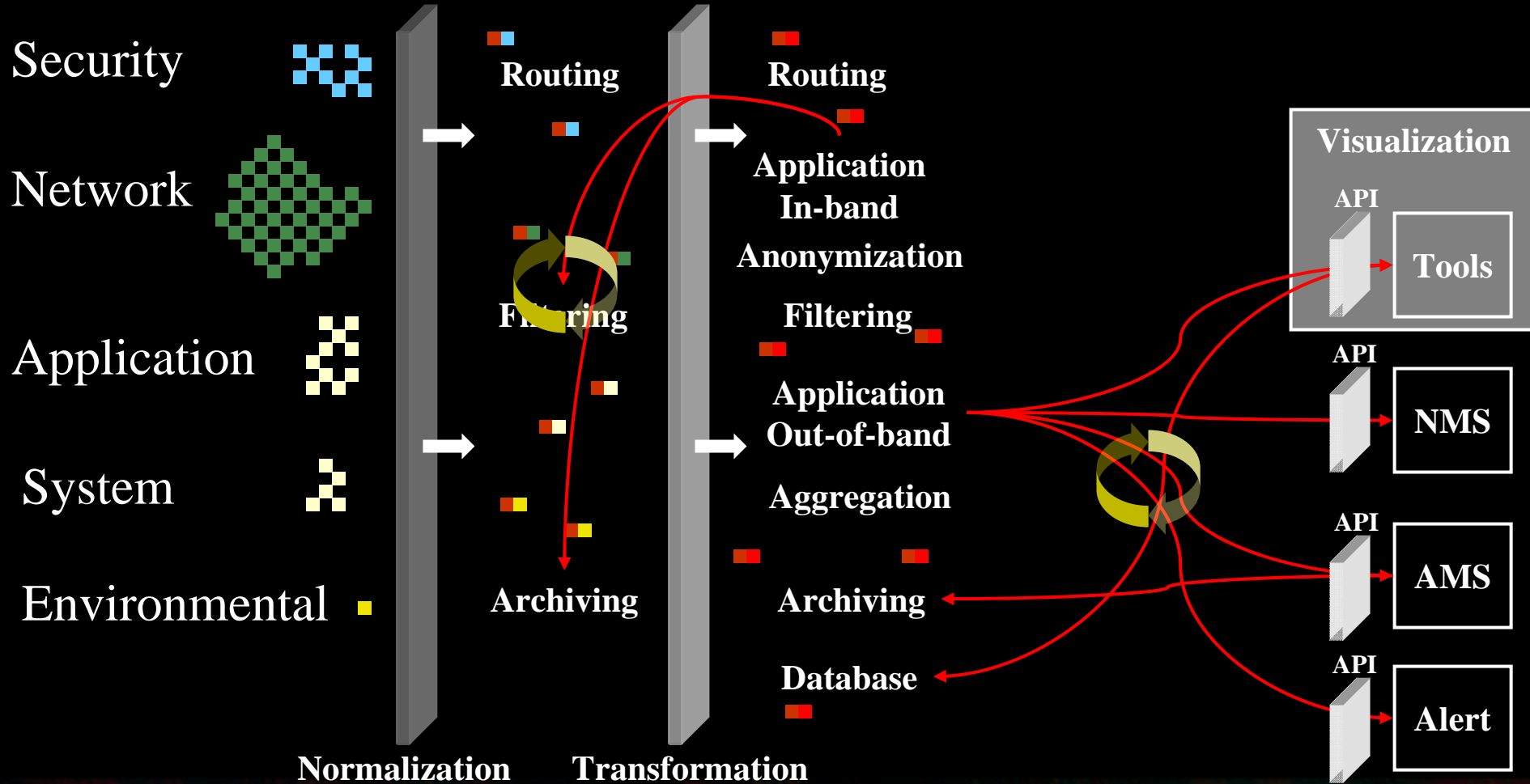


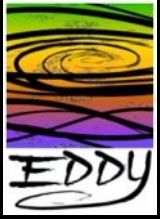
# Event Data Types

- Network
  - Fault, performance, flow, etc.
- Security
  - IDS, auditing, conformance, etc.
- System
  - Hardware faults, RAID/SAN performance, etc.
- Application
  - Fault, performance, auditing, profiling, etc.
- Environmental
  - Building sensors (lighting, sound, temperature, power, actuators, pressure, stress, etc.)

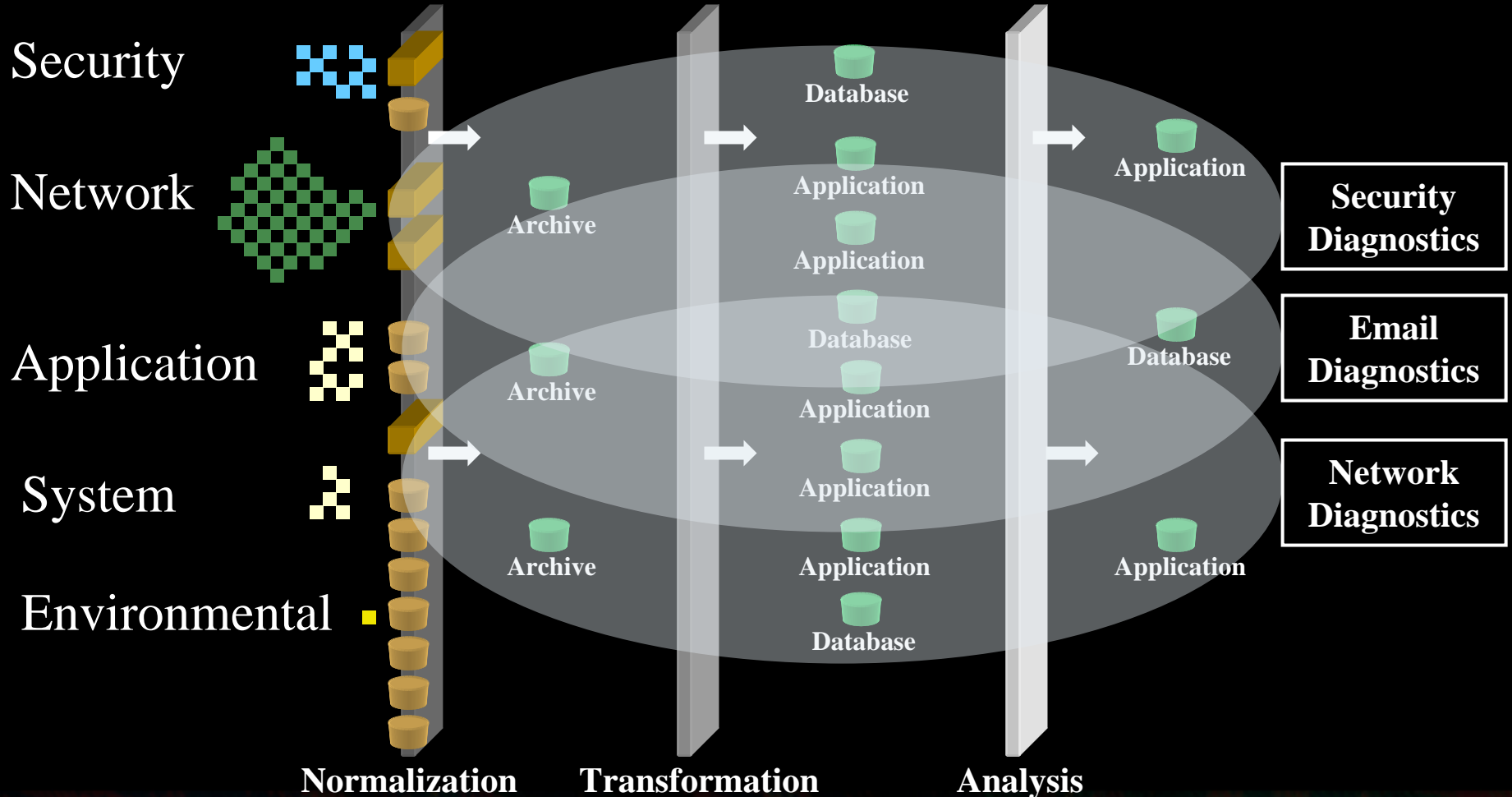


# EDDY Event Evolution





# EDDY Cluster Functionality



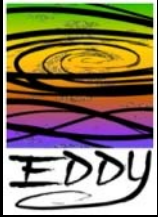
Edge Nodes



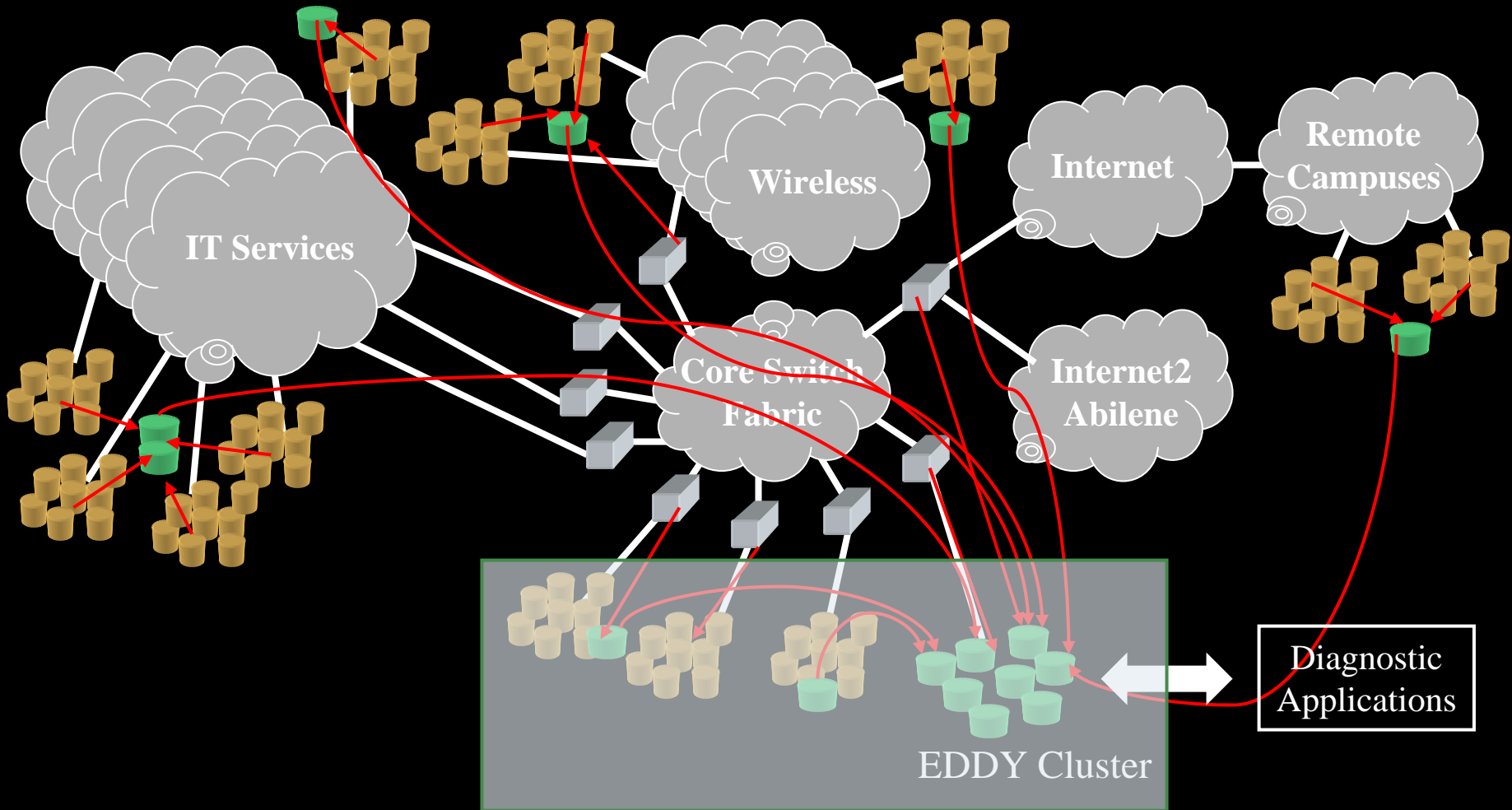
Backplane Nodes



Carnegie Mellon



# Edge and Core Events



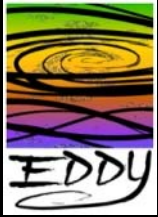
Edge Hosts



Backplane Hosts

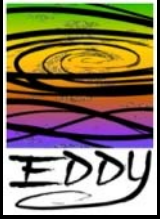


Carnegie Mellon



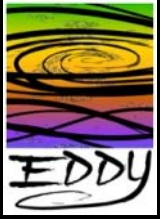
# The Scale Issue

- Network events
  - Flows on high value network points aggregation (e.g. core routers/switches egress (commodity and Internet2, wireless))
    - Core routers/switches is at least 5000 flow events/sec average
  - State and summary (e.g. SNMP, RMON, etc.)
- Application events
  - Email generates about 250 events/sec average



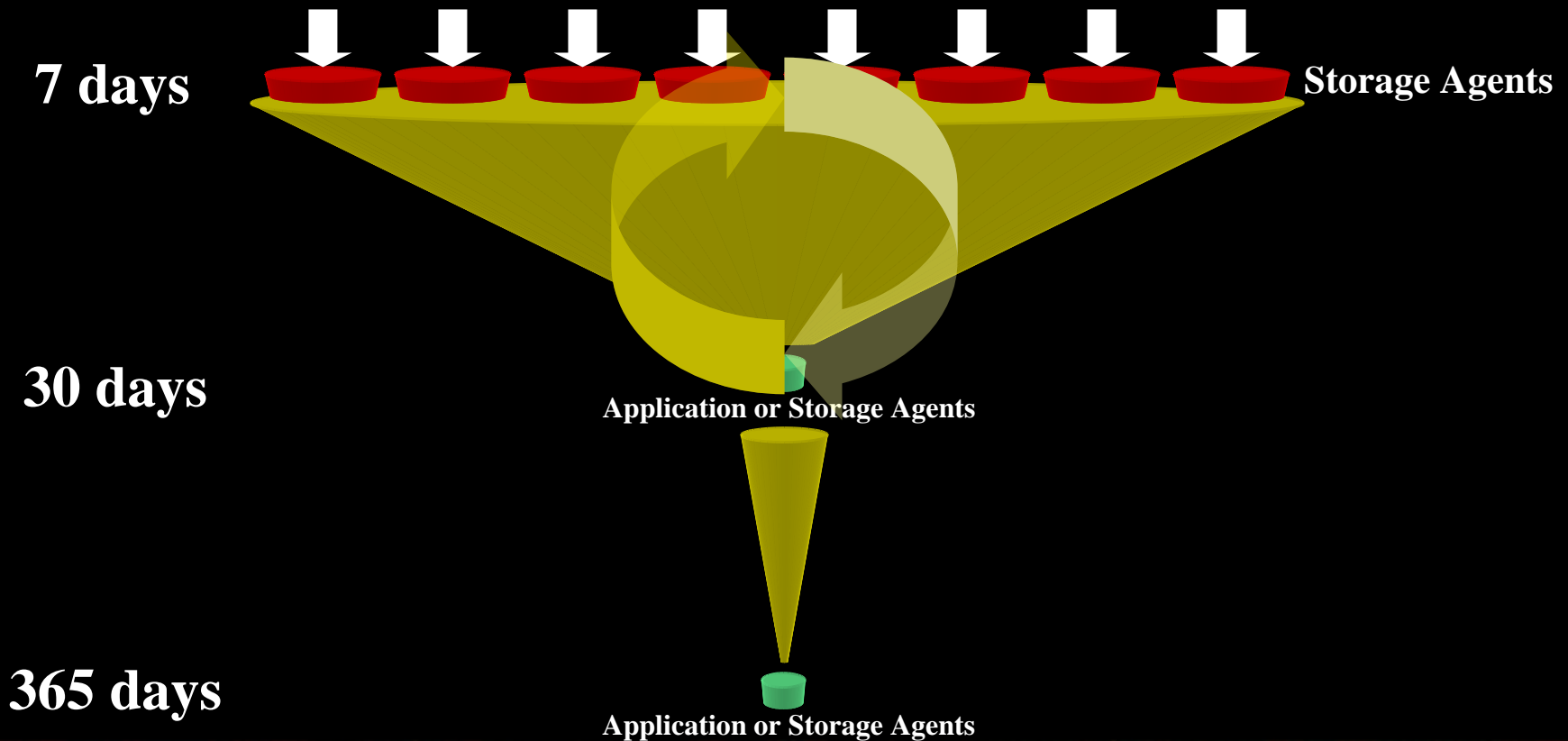
# The Scale Solution

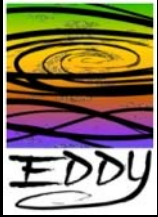
- Scalable store and forward
  - Select/Project forward only what is needed
  - Query back to get data that you don't have
  - Only cook data that you need
- Data lifecycle



# Example of the Approach to the Scale Issue

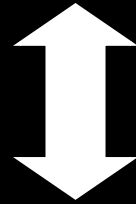
Each event stream is 5k/sec





# Diagnostic Data Lifecycle

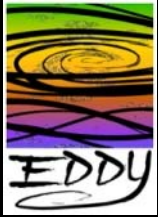
Access



Collection → Anonymize/  
Filter → Aggregation → Archive → Scour

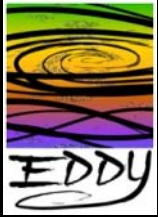
Policy

Time



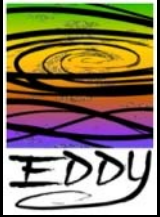
# Outline

- Motivation
  - The EDDY approach
  - Architecture and technical challenges
- **EDDY enabled applications**
  - **Current and planned efforts**
- Release 0.5
- Current Activities



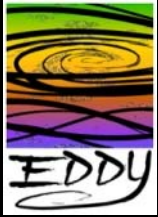
# Rapid Enabling of Diagnostic Applications

- Enable the forensic process in a broad manner
- Feeding existing analysis tools/systems to enhance their functionality
- Expose event information to enable new visualizations to represent real-time and historical events
- Feeding research with a vast and diverse array of data



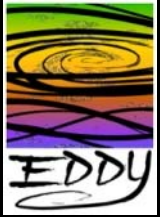
# EDDY Enabled Devices

- Workstation and servers
- Network devices (routers and switches)
- Security devices (firewalls and IDS)
- Embedded EDDY
  - Environmental devices (premises control/monitoring)
  - Transportation (automotive, etc.)
  - Robotics



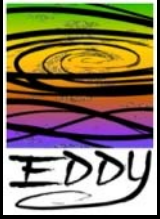
# Present Focused Problem Domains

- Security
  - taxonomic risk analysis
  - exposing events to enhance the forensic process
- Middleware
  - Email
- Environmental
  - energy conservation
  - human factors
  - forensic process of building faults



# Consumers of EDDY Environmental Events

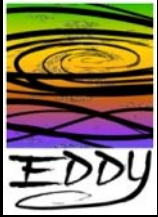
- Power Management
- HVAC Management
- Security Management
- Utilities
- Scientists
  - Analysis of building designs, lifecycles and their occupants



# Network and Security Diagnostics

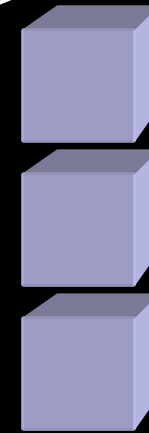
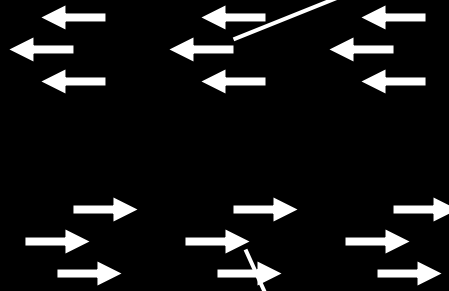
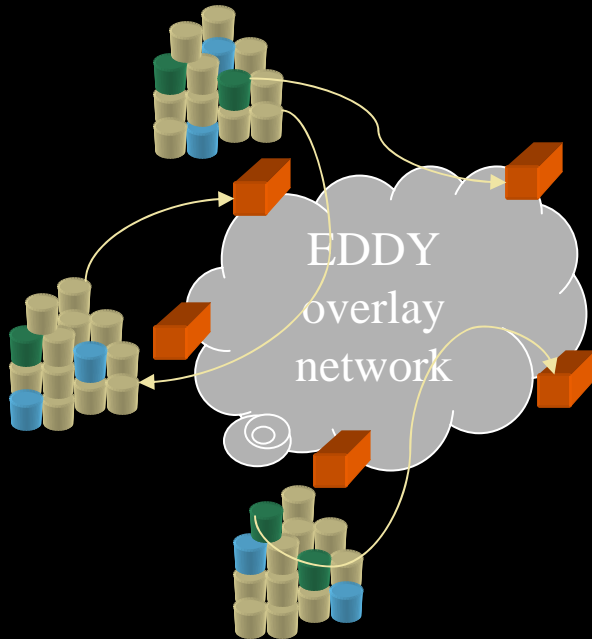
## Questions answered

- Who is using the most bandwidth through the egress?
- What network flows were associated with this Snort event?
- What systems have the worm and what is the rate are they infecting others?
- Why is the application running so slow?



# Feeding Existing Apps

- Adjusting EDDY agent aggregation flows
- Queries to EDDY event repositories







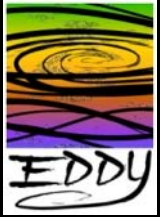
Network Management  
(TopTalkers)

Security Management  
(Project Dragnet)

Network Management  
(flow forensics)

- High order events from EDDY applications
- Real-time event streams (4-5K events/sec)
- Query responses from EDDY event repositories

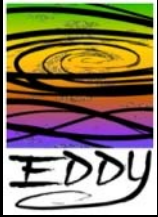
-  EDDY application agents
-  Network Events
-  Application/System Events
-  Security Events



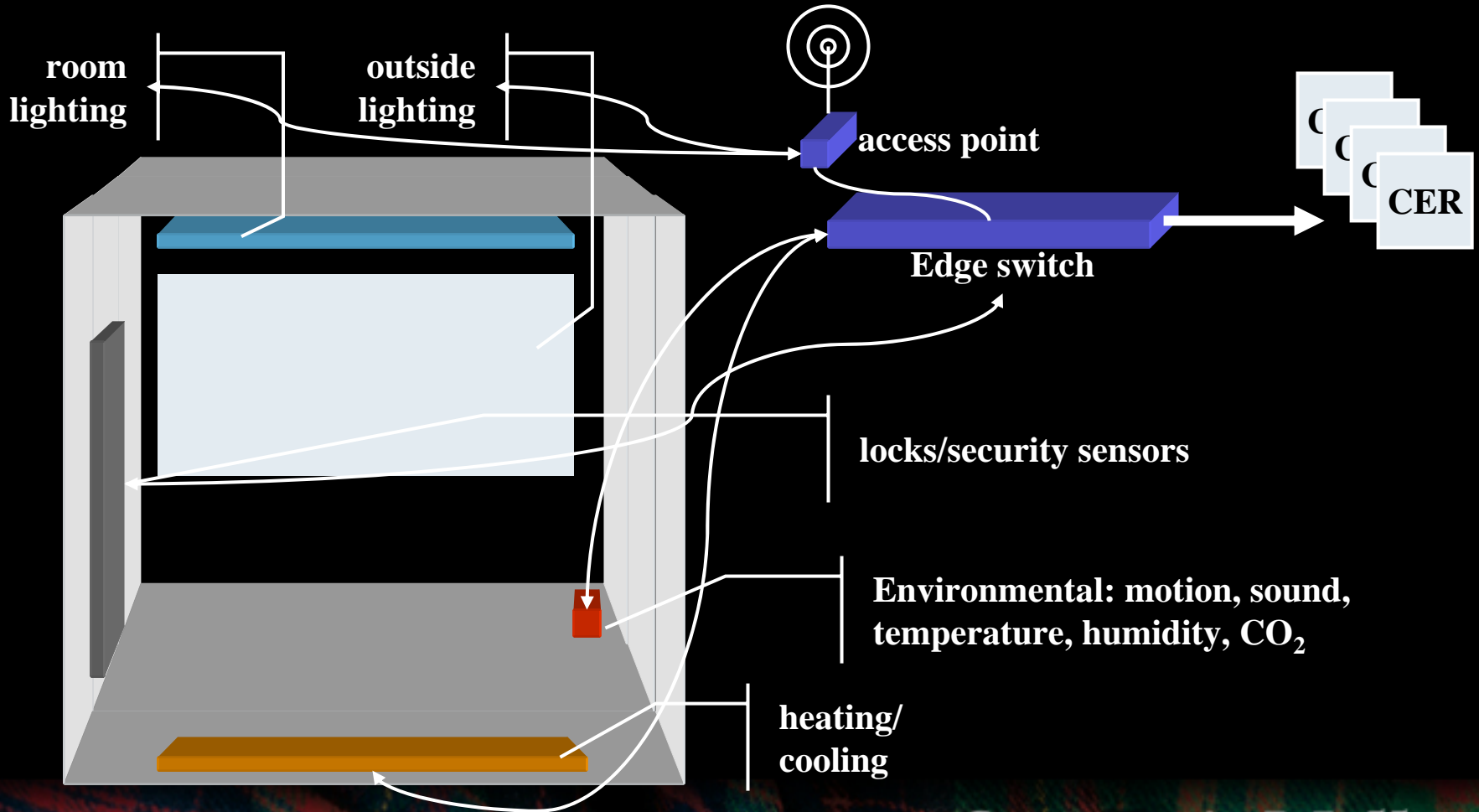
# Environmental Diagnostics

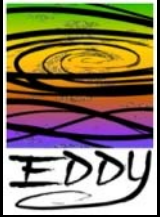
## Questions answered

- Why is it so cold on the north side of the building?
- Can we reduce the heat after 7pm?
- How much energy is the new HVAC system using?
- Was anyone in my office last night and read my Email?

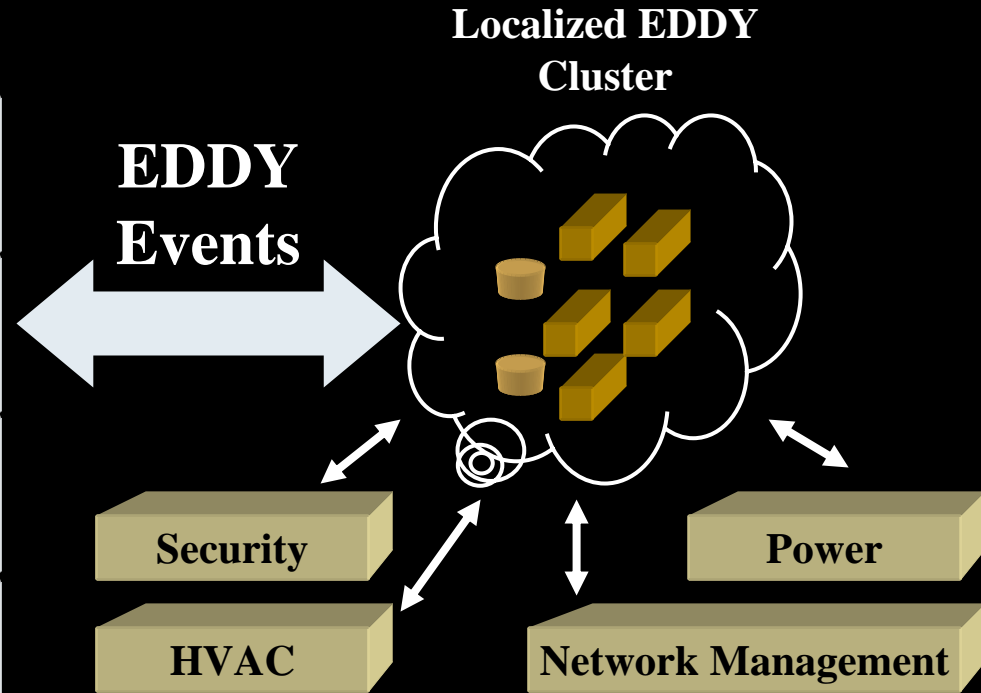
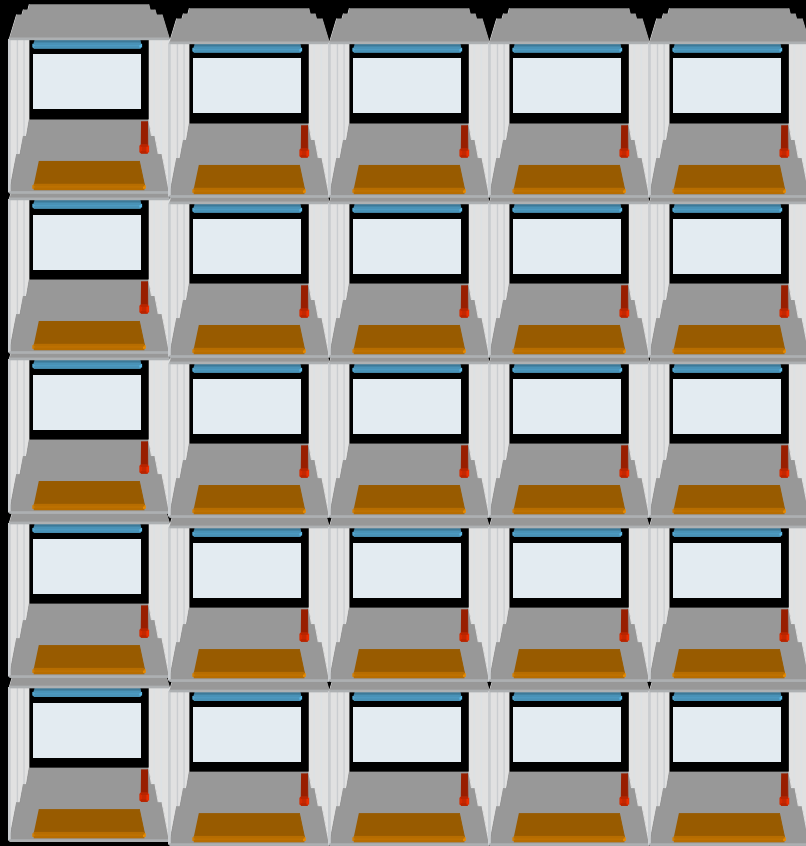


# Environmental Events

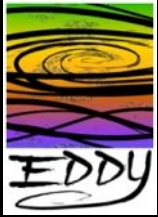




# Environmental Events



**EDDY Enabled Applications  
(at campus BOC)**



# Email Diagnostics

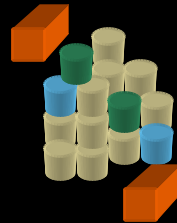
CER(Email): rejected because...





CER(Email): Spam Queue is...

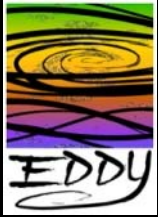
CER(Email): received



Email



-  EDDY analysis agents
-  MX hosts/Spam engines
-  Frontend Cyrus (IMAP)
-  Backend Cyrus (storage)



# Email Diagnostics

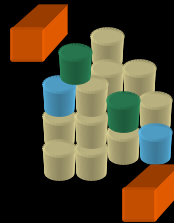
CER(Email): rejected because...

CER(Email): Spam Queue is...

CER(Email): received



Email



CER(Email): mail-thread

CER(Email): received

CER(Email): Spam Queue is...

CER(Email): rejected because...



EDDY analysis agents



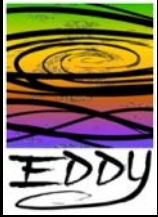
MX hosts/Spam engines



Frontend Cyrus (IMAP)



Backend Cyrus (storage)



# Email Diagnostics

## Questions answered...

### Message delivery:

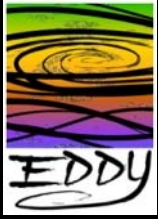
- Why was this message delayed?
- Why was it delayed for me but not you?
- Why wasn't this message received?
- Why was this message received by X but not Y?
- Why was this message bounced?
- Why did I receive this message when it wasn't directed to me?

### Host information

- Why is this machine slow/slower than the rest?
- Are slow LDAP lookups resulting in cascading failures?
- Is there a DNS issue?

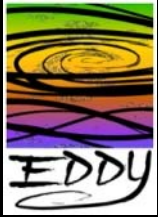
### Reporting/analysis

- How many messages did I process?
- What is my projected growth rate?
- Where is my next bottleneck?



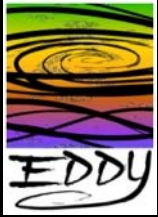
# Outline

- Motivation
  - The EDDY approach
  - Architecture and technical challenges
- EDDY enabled applications
  - Current and planned efforts
- **Release 0.5**
- Current Activities



# EDDY v0.5 Software Distribution Contents

- EDDY Agent Framework
  - Every EDDY Agent extend from this code which provides core EDDY capabilities to all Agents.
- TLS Scripts
  - Used for generating and signing certificates and putting them into keystores used by Java.
- Sample EDDY Agents
  - 18 Normalization, Transformation & Display agents.
  - Agents Produce Raw, Cooked & Analysis CERs.
  - Template Agents used as starting blocks for Java and Perl Agent development.
- Agent Manager
  - Management software to start and stop EDDY agents on a host.
- Miscellaneous Files
  - Files that are used with the sample code, documents, CER schemas, etc.



# EDDY Agent Framework Design

## Common Capabilities

The EDDY Agent Framework provides functionality for tasks all EDDY Agents perform such as transport, filtering and routing and thread management

## Ease of Use

The Java object-oriented of the EDDY Agent Framework makes it easy for developers to get up to speed and produce EDDY Agents quickly

## Robustness

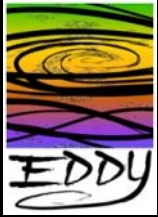
The EDDY Agent Framework code was design and implemented to work reliability under stress and to anticipate and handle error conditions well

## Scalability

EDDY Agent Framework based Agents can handle marshalling and unmarshalling of CERs at rates of 15,000+ per second commodity hardware

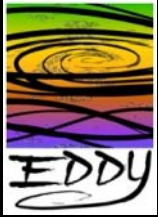
## Flexibility

The EDDY Agent Framework is flexible enough to support development of any type of EDDY Agent while providing capabilities that all Agents require



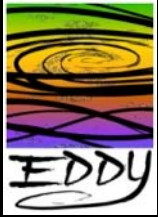
# Developer Friendly Agent Framework

- Agents inherit filtering, routing and TLS transport capabilities from Agent Framework.
- EDDY Agents extend from Agent Framework to add unique functionality.
- Object-oriented model used for “overloading” EDDY Agent Framework methods.
- Simple, but elegant model to build EDDY Agents from the ground-up quickly.
- Lots of sample code that demonstrate functional EDDY Agents.
- EDDY Developer & Deployment Guide steps through building a real Normalization, Transformation and Display agent



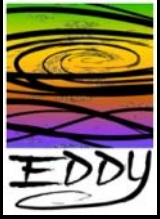
# Outline

- Motivation
  - The EDDY approach
  - Architecture and technical challenges
- EDDY enabled applications
  - Current and planned efforts
- Release 0.5
- **Current Activities**



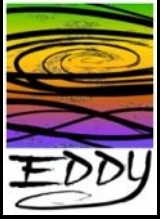
# What EDDY is

- Architecture for cross domain diagnostics
- An enabling technology that provides,
  - Event ledger
  - Dissemination and correlation infrastructure,
    - Afford research access to event data (anonymized)
  - A development platform for diagnostic research
    - Domain specific
    - Domain agnostic



# What EDDY is not

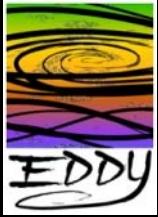
- A system/network/application/security management platform
- The analysis engine, it enables the analysis to happen with domain expertise



# Unleashing the Potential

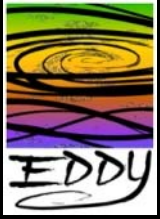
Exposing an unprecedented wealth of diagnostic information for

- Enabling new and enhancing existing diagnostic and security applications
- Environmental analysis and energy conservation
- Visualizing events
- Security forensics
- Modeling new policy configurations to assess their impact on daily operations
- Analysis, validation and troubleshooting of distributed composite applications



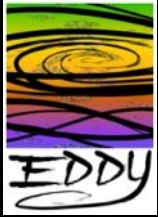
# Next Generation Tools

- Inter-domain event correlation, e.g.
  - Network, application, system and security event domains combined
  - Performance impacts across layers
- Domain agnostic analytics
- True end-to-end accountability of transactions
- Establish rich service profiles based on events



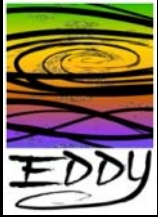
# Enabling Campus Members

- Data for extended research
  - A platform to discover new diagnostic application methods
  - Providing a “Petri-dish” for researchers to gain access to a wide range events
- Enterprise diagnostics
  - Within CMU (e.g. Computing Services)
  - Other federated applications



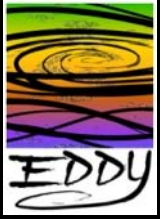
# Enabling Campus Members Cont.

- Richness of data...whatever you imagine.
  - Security (flow data, IDS, cameras, log files, etc.)
  - Distributed systems
    - Applications (Outlook, calendar, etc.)
    - Middleware (DNS, Kerberos, DNS, sendmail, etc.)
    - Computational devices (RAID, disk, memory, embedded, etc.)
    - Network (flow data, SNMP, RMON, etc.)
  - Sensors (seismic, environmental, robotics)
- The ability to access data from remote locations (with policy restrictions of course) and correlate it



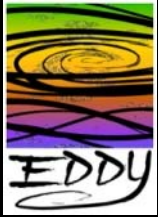
# Campus Adopters

- Architecture School/Intelligent workplace
  - Environmental monitoring and control
  - Prototyping the next generation of environmental diagnostics
- CS/Cylab – security research, real time flow events from CMU commodity Internet
  - Dragnet – network flow event security analysis in real-time up to 10K events/sec
  - Abilene (near real-time) to follow



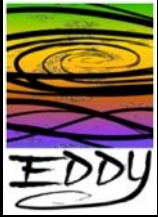
# Campus Adopters Cont.

- Computing Services ISAM/Security Office
  - Consolidation of application log files, fault analysis (top network talker events and email infrastructure)
  - Conduit for reporting and high level event consumption



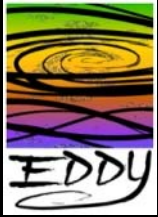
# Ongoing Efforts

- **Architecture:** A solution for integrating the diagnosis of distributed network and systems
- **Standards:** Defining the next generation of event auditing (working with IBM and others)
- **Open Source Prototype:** An efficient event dissemination platform that can be installed on the end system or within network devices
- **Center for Diagnostic Research:** Help in thinking about correlating security, application and network events and other applications



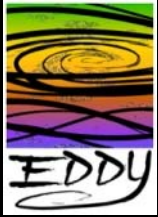
# Status

- Development
  - Initial release (Munster 0.5) targeted at developers - 4/1/06
- Outreach
  - Involving others in the development process (CMU and at Duke)
  - Expand to other use cases external to CMU
  - Presently working with industry leaders on proposed standards and methods
- Support
  - Sponsored by the National Science Foundation under the NSF Middleware Initiative - Grant No. ANI-0330626
  - Soliciting collaboration partners in both industry and government



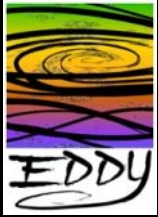
# Future Development Directions

- Simplify the injection of new events
  - Add new types of CER
  - Toolkits for other languages (e.g. Perl)
- Mature the CER
  - New Features: dynamic, hints
  - Formalization of the structure of userTag field
  - Standardize: Working with IBM and WSDM, Cisco on transport
- Develop the query and control channel
- Add new high value agents (e.g. storage, discovery)



# Want to Learn More?

- Web site
  - [www.cmu.edu/eddy](http://www.cmu.edu/eddy)
- Mailing list
  - [Eddy-info@lists.andrew.cmu.edu](mailto:Eddy-info@lists.andrew.cmu.edu)



# The EDDY Initiative

Progress to Date

**Chas DiFatta (chas@cmu.edu)**

**Mark Poepping (poepping@cmu.edu)**

**Carnegie Mellon**