

A Brief Overview of the CyDAT Diagnostic Effort

CyDAT - Cyber-center for
Diagnostics, Analytics and Telemetry

Chas DiFatta (chas@cmu.edu)
Mark Poepping (poepping@cmu.edu)

Diagnostics...?

You discover your car has a flat tire...

- You fix it you move on

It's flat again a week later...

- Valve problem?
- Nails in the driveway?
- Neighbor kid?

Can you check all failure possibilities?

- Might help if you knew when air started leaking

Cars... Computers...

You discover your Sendmail daemon crashed...

- You restart it and you move on

It crashes again a day later...

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

Why Diagnostics?

- Things break, in complicated, partial ways – and it matters
- Systems built to ‘get it working’, not to be ‘fixed’
 - Meter/maintain/fix after installation?
 - The maintainer learns how... but it’s a struggle
- Software reuse and layered infrastructures create dynamic dependencies
 - Diagnostic data may not be available at all
 - Certainly doesn’t follow service path
 - Minimally ‘out of band’, often ‘out of question’
- Service Plane + Management Plane + *Diagnostic Plane*

Problem Statement

- Not enough...
 - cooperation across complex, layered, networked objects
 - access to existing data relevant to diagnoses
 - awareness of leading indicators how to apply data in practice
 - information in the data currently available
- But already too many...
 - data formats to reasonably understand or correlate
 - data sources and records to easily manage
- And it's getting harder...
 - Software infrastructures (Middleware, Federation, SOA, ESB)
 - IT-enabling *everything*
 - Financial and Security audits are expanding
- Data mining in real time
 - Leverage data flow to augment store+search
 - Needle in a needle-stack

Improve Diagnostic *Practice*

- Make routine scenarios automatic
- Bring more data to new situations
 - Active probes plus passive activity collection
- Enable cross-domain practice
 - IT Security
 - IT Audit
- Augment domain-specific practice
 - In situ integration of Network, Middleboxes, Systems, Middlewares, Applications, Reporting
- Document experience to improve inputs

Structuring the Problem

[An Architecture for Diagnostic Infrastructure]

Sensing Technology

- State, transaction info, whatever...the ability to collect anything

Diagnostic Data Orchestration

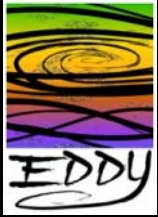
- Data acquisition/normalization/transport, getting the:
 - Instrumentation data you want
 - In the format that you need
 - Where you want it

Diagnostic Information

- Generic translation and statistical methods
- Simple event correlation, visualization, longitudinal pattern analysis
- Data Lifecycle (must be policy driven)

Domain-specific Diagnostic Analytics

- Detailed analyses, situational diagnosis, specialized UI's
- Significant automation of the domain and implementation autonomies



EDDY Capabilities

[Orchestrate Data and Create Generic Information]

Enable correlation

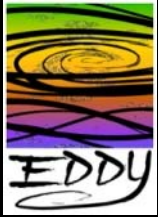
- Common Event Record (CER) – a way to format event information to make it easier to process

Provide transport

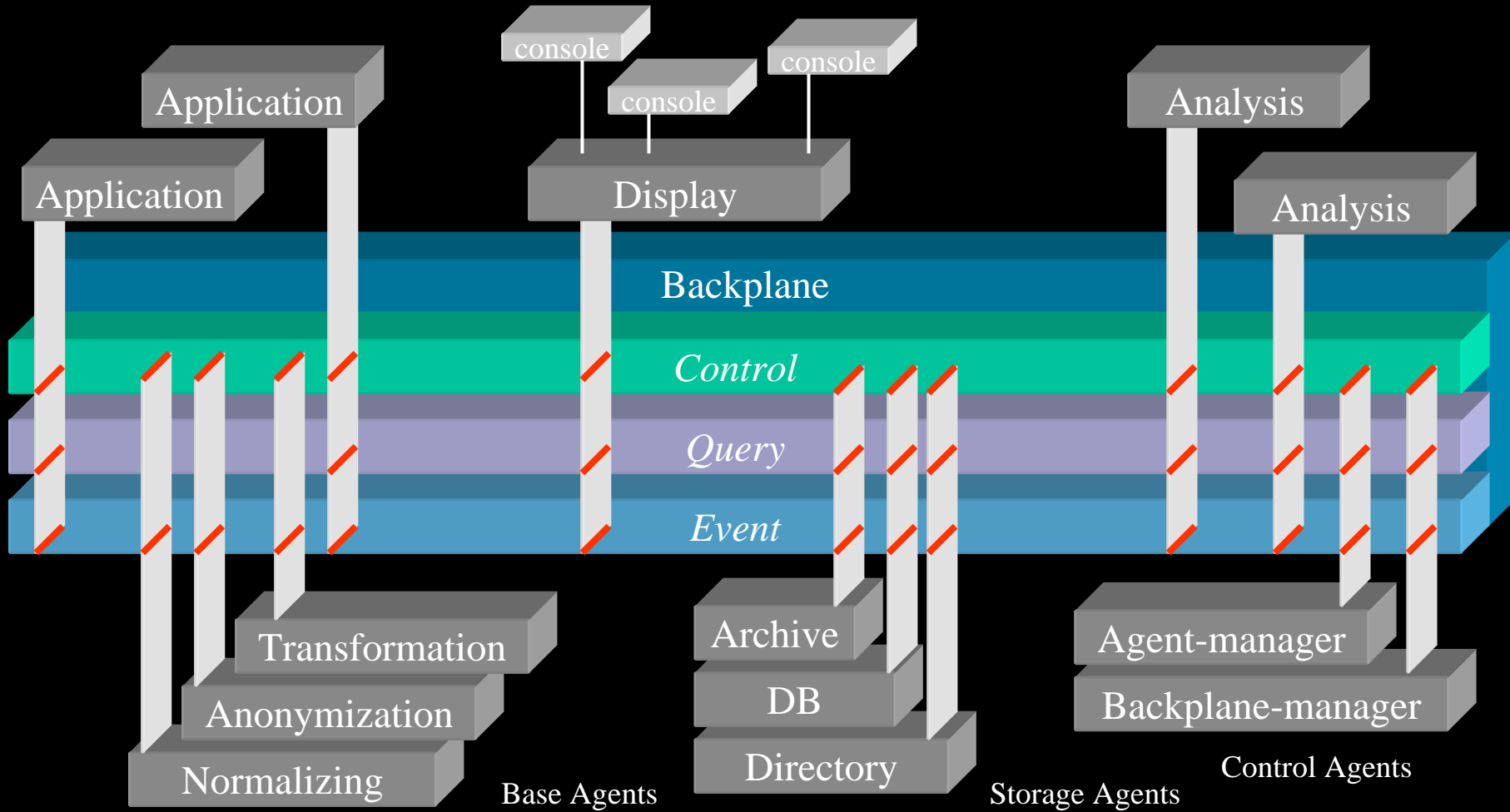
- Diagnostic Backplane – a way to move CERs around to make it easier to automate processing

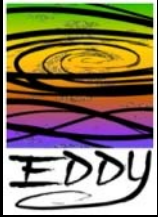
Some simple event orchestration methods

- Normalize, transform, visualize, store, anonymize



EDDY Backplane





EDDY Status

- Development
 - Initial release (Munster 0.5) targeted at developers - 4/1/06
 - EDDY Agent Framework, Sample EDDY Agents, Agent Manager
 - Supplemental release (Sushi 0.5.1) targeted at network managers -8/16/06
 - Framework enhancements, Normalizers, TopNetworkTalker Application
 - Supplemental release (Murphy 0.5.2) targeted at developers – 10/12/06
 - Improved internal diagnostics
 - Supplemental release (Murphy 0.5.2.1) targeted at developers – 4/5/07
 - Dynamic agent filtering
 - Supplemental release (0.6) targeted at Email and security diagnosticians
 - Normalizers, Email and security application, additional performance increases
- Outreach
 - Involving others in the development process
 - Expand to other use cases external to CMU
 - Working with industry leaders on proposed standards and methods
- Support
 - Initially sponsored by the National Science Foundation under the NSF
 - IBM sponsored open diagnostic effort – call to action of industry leaders
 - Compute hardware support from SUN
 - Other industry conversations in progress

Overall Vision

World Leader for *Awareness* of the Physical and Virtual Infrastructures on Campus

- Most sensed and aware environment
 - Building management, power, structure, environmental
 - Constituent activity, health, location
 - Information Technology infrastructure/support services
- Most diagnosable environment
 - Create an overall diagnostic infrastructure
 - Build an event repository for research
 - Domain-specific and domain-agnostic analytics

CyDAT Proposal


Cyber-center for Diagnostics, Analytics, and Telemetry

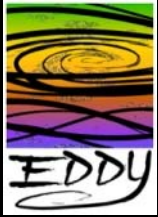
- Create critical mass and center of gravity to invent and evolve an infrastructure for orchestration, management, and analysis of diagnostic data
- Center is agnostic of diagnostic domain
- Support and interact with domain-specific efforts
 - Informed by use in multiple application domains
 - Experience with engineering and research use
 - IT-enabled, not IT-specific

CyDAT Strategy

- Clarify the Architecture
 - Act as a neutral party to promote a non-biased view on an common event infrastructure with key influential members of industry
- Build a Toolkit
 - Define and build a rich reference implementation platform for event dissemination
- Use it for real research and engineering
 - Leverage real use across campus
 - Create a data environment and analysis platform to support domain-specific and inter-domain experimentation

CyDAT Organization

- **Architecture and Standards**
 - Design and define specifics for the IT Diagnostic Plane
 - Commercial support and development (IBM and others)
- **Open Source Prototype** 
 - Create a solid reference implementation for experimentation with the Diagnostic Plane
- **Diagnostic Observatory**
 - Leverage a large-scale event facility at Carnegie Mellon for engineering and research collaboration on real data
 - Computing Services provides data, needs engineering analyses
 - Facilitate data export to other researchers
 - Research on structure and behavior of the Diagnostic facility

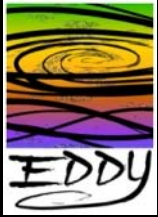


Development and Engineering Goals

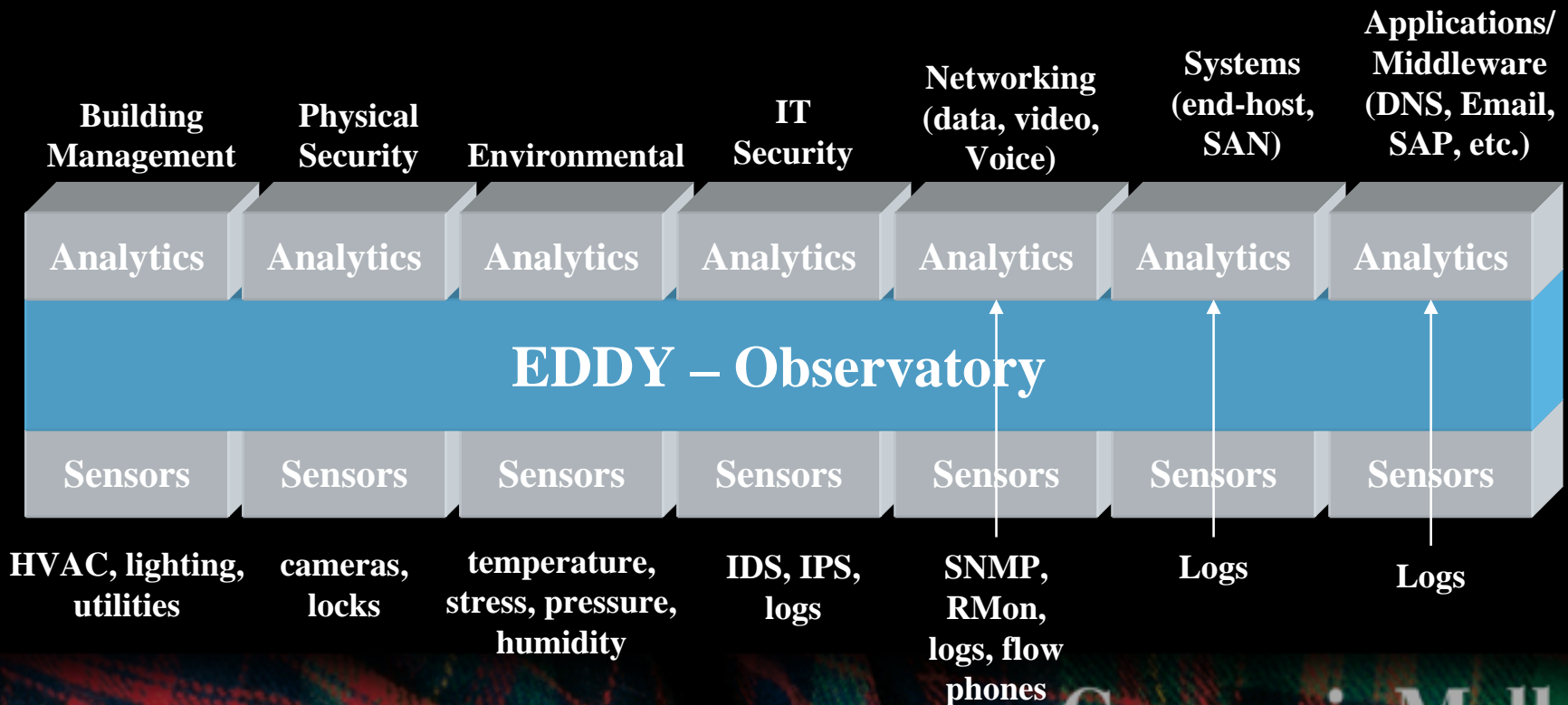
- Architecture - Identify and assemble technology leadership
 - Refine problem space and strategic requirements
 - Identify domains of interest with existing, emerging, or required standards
- Standards
 - Define next generation standard event format based consortium work
 - Define telemetry interfaces and their core features
- Open source reference implementation
 - Implement feature roadmap (next)

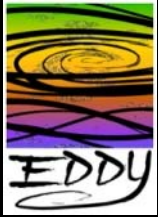
Observatory Goals

- Initiate an effort to establish an event observatory at Carnegie Mellon that coordinates and connects domain-based distributed event infrastructures
 - Network Flows
 - IT Infrastructure information
 - Physical sensors
 - Other data of interest



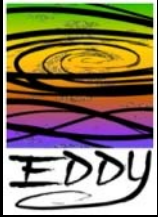
EDDY and the CyDAT Event Observatory





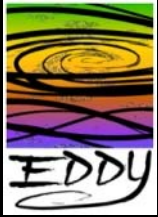
Demos

- Email
 - Focused on the email administrator and help desk
 - Four basic questions (to/from/mesg ID, host)
 - Roll your own query
- Top Network Talkers
 - Focused on the network/security administrators
 - Perspective from the flow, byte and



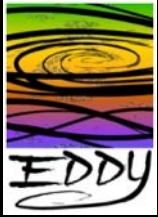
Lessons Learned in Developing Email Diagnostic Application

- Traditional Syslog can't keep up
- Using DB as an initial store for building email “blob” is a lose
 - Inserts to DB take too long
 - Searching complexity



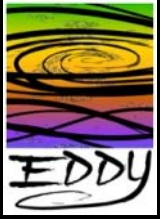
Lessons Learned in Developing Network Diagnostic Application

- Simple statistic methods can give real value quickly
- *Constant* desire for the diagnostician to combine data from other sources
- Real-time is great, but historical is always wanted, both for trending and snapshots of anomalies



CyDAT Status

- Established in Carnegie Mellon CyLAB
 - Co-sponsorship: CyLAB, Computing Services
 - www.cylab.cmu.edu
- Foster Additional Collaboration
 - IBM/CMU on lead on seeding diagnostic collaborative
 - Build on Engineering and Research Activities
 - Using classic diagnostic tools/toolkits
 - Integrating with EDDY
 - Research on EDDY futures



Campus interactions...

ISO - Traffic analysis

- Security diagnostic applications

ISAM – Email message transport

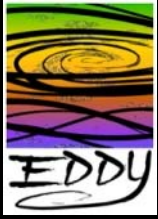
- Email diagnostic applications

Computer Science – Dragnet (Sekar, Zhang)

- Forensic analysis and auditing methods in real-time.

School of Architecture – Intelligent Workplace

– Sensing the environment



Campus interactions...

CyLab – (Kim, Reiter, Turner, Wing, Yen)

- Network telemetry, real-time security analysis, BOTNet detection

Civil Engineering – CenSCIR

[Center for Sensed Critical infrastructure and Research]

- Large scale orchestration of environmental events from externally and internally located sensors

Other discussions – PDL, CERT, ECE

[Parallel Data Lab, Computer Emergency Response Team, Electrical and Computer Engineering]

- Data center large scale computing applications
- Security applications
- Distributed systems diagnosis

Questions/Comments

EDDY Effort: www.cmu.edu/eddy

A Brief Overview of the CyDAT Diagnostic Effort

CyDAT - Cyber-center for
Diagnostics, Analytics and Telemetry

Chas DiFatta (chas@cmu.edu)

Mark Poepping (poepping@cmu.edu)