

Unified Federated Diagnostics BoF

Leveraging multiple sources of
diagnostic data

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

Outline

- Status of CyDAT/EDDY Efforts
 - Collaborative
 - Observatory
 - October Software Release
- EDDY feature roadmap
- Discussion on future diagnostic tool and method requirements

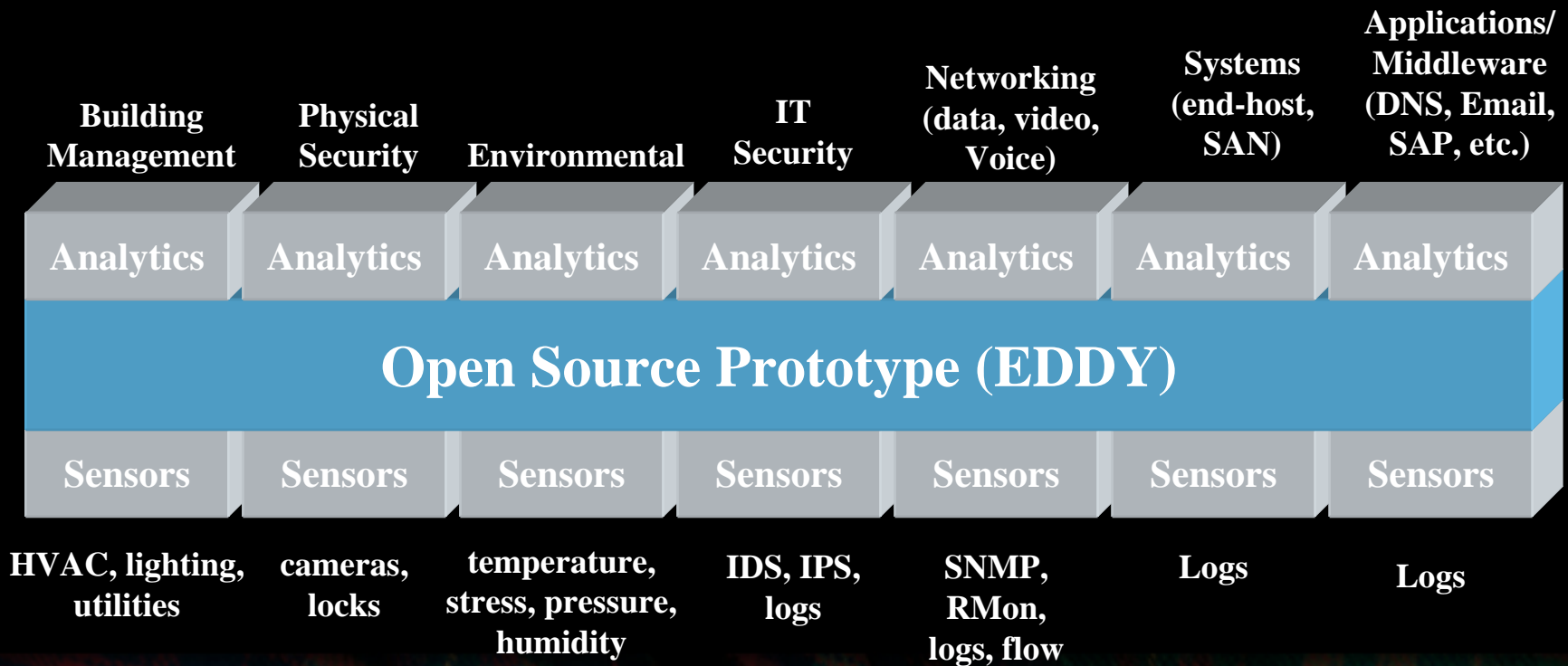
Cyber-center for Diagnostics Analytics and Telemetry (CyDAT)

- Foster Additional Collaborators
 - IBM/CMU on lead
 - Build on Research Activities
 - Research Using EDDY
 - Research/engineering on EDDY
- Research Collaboration within HiED

CyDAT Observatory Goals

- Establish Facility at Carnegie Mellon
 - Coordinate/connect disparate events
 - IT Infrastructure information
 - Network Flows; Application, Middleware logs
 - Physical sensors
 - Temperature, power consumption, water flow, traffic
 - Other data as use cases (and research) guide
 - Support for experimental analytics/autonomics
 - Sensitive to and supportive of security/privacy

CyDAT Observatory



Observatory Data Operations

Security and network diagnostics (16K events/sec)

- Network flow data (both from egress and core networks)
- Top talkers (packets, flows, bytes)
- Various analytics: Botnet, Topnete, FlowView, etc.)

Email (500 events/sec)

- Log data from all MX, Spam, Cyrus hosts
- Creating an Email flow record
- Analytics: Email View

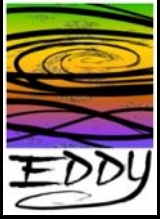
Observatory Data Operations Cont.

Environmental / Building management

- Various sensor data (Firefly, SunSpots, etc.)
- Analytics: in development

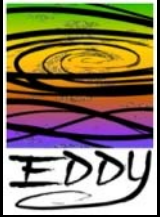
System management

- Lots of logs: Mon, syslog, Shibboleth, SAN
- Analytics: ShibView, SANview (in development)



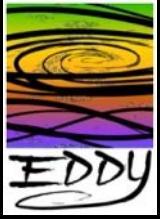
What EDDY is

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



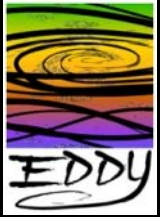
EDDY Release 5.2.2

- What's new?
 - CER Factory and Perl client for injecting log data
 - Eclipse as a development environment
 - Easy deployment
 - RPM
 - Tarball with Ant files



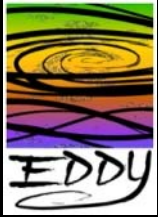
CER Factory Concept

- Facility to easily create CERs from a limited set of event attributes
- Simplify using the backplane for non-Java developers
- Reduce the barrier to adoption for embedded system developers
- Provide a wide array of transport options (SSL, HTTP, SOAP, native sockets, etc.)

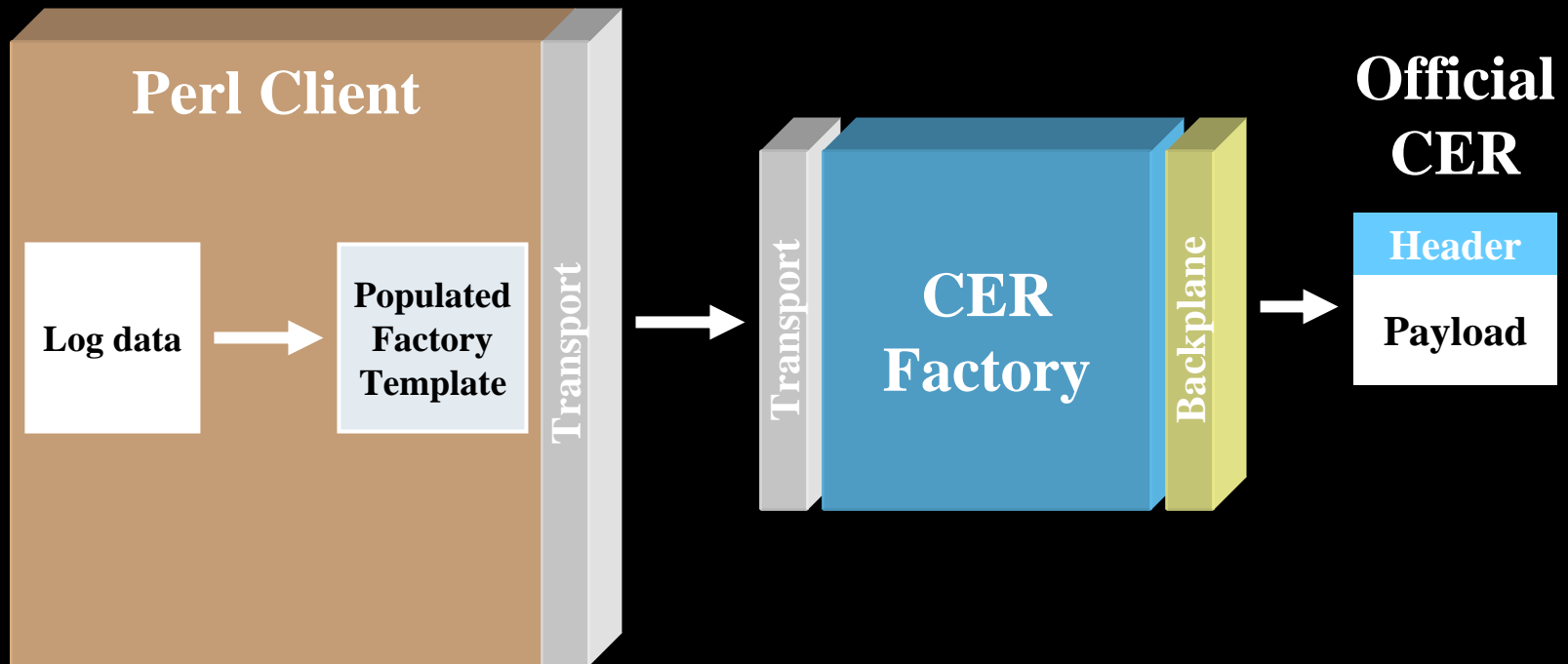


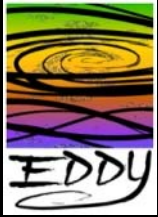
What EDDY (End-to-end Diagnostic DiscoverY) is

- 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



Basic Architecture

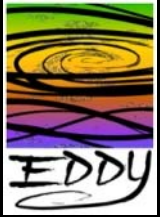




CER Factory Template Example

Shib event as a Factory Template:

```
<cerFactory>
  <eventInfo.oid>1021</eventInfo.oid>
  <eventInfo.userTag>
    <key>filename</key>
    <value>/afs/andrew.cmu.edu/usr23/work/eddy/cerfactory/transaction.log</value>
    <key>generationProcess</key>
    <value>transaction</value>
  </eventInfo.userTag>
  <dataPayload.payloadType>2</dataPayload.payloadType>
  <dataPayload.payload>
    <shib>
      <shib-1.0.0>
        <subEventID>123</subEventID>
        <id>b278f62db622adf56b642eaeefb83bc8</id>
        <logData>
          Caching the following attributes after AAP applied for
          session (ID: _b278f62db622adf56b642eaeefb83bc8) on (applicationId: default) for
          principal from (IdP: https://idp.testshib.org/shibboleth/testshib/idp) {
          urn:mace:dir:attribute-def:eduPersonScopedAffiliation (1 values)
        </logData>
      </shib-1.0.0>
    </shib>
  </dataPayload.payload>
</cerFactory>
```



EDDY Feature Roadmap

- Internal diagnostics
 - Internal agent error
 - Periodic summaries of agent operation
- Storage agent
 - High performance (in memory)
 - High capacity (on disk)
- Normalizers
 - NetFlow V9, DNS, SAN, Snort, etc.
- Visualization!!!

Discussion

- *Attaching generic tools to the sink*
 - *Open-source stuff*
 - *Easy stuff*
 - *Fancy stuff*
- *Intermediate log processing tools*
- *What is your pain?*

Want to Learn More?

- Web sites
 - www.cmu.edu/eddy
 - www.cmu.edu/eddy/cydat
- PIs
 - Chas DiFatta (chas@cmu.edu)
 - Mark Poepping (poepping@cmu.edu)

Unified Federated Diagnostics BoF

Leveraging multiple sources of
diagnostic data

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