Distributing weather data via multipoint layer-2 paths using DYNES

Malathi Veeraraghavan
University of Virginia (UVA)
mvee@virginia.edu

Steve Emmerson
Univ. Corp. of Atmospheric Research (UCAR)
emmerson@ucar.edu

April 7, 2014
Internet2 Global Summit

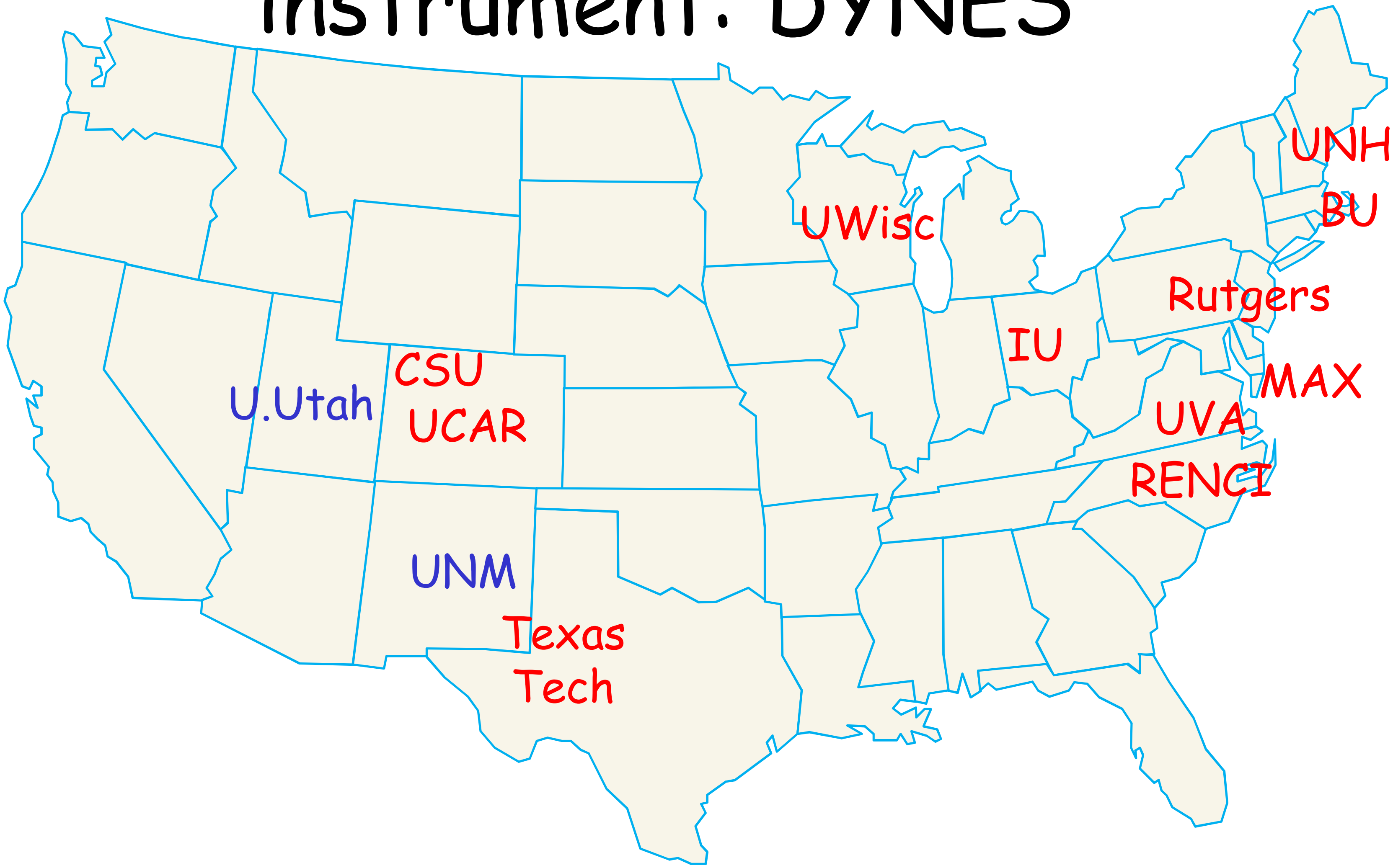
- Science application: Atmospheric science
- Technology to support reliable multicast
- Integration and test plan

Supported by NSF CC-NIE grant: ACI-1340910

Acknowledgment

- Collaborators: Steve Decker and Ivan Seskar, Rutgers Univ., Jerry Robaidek and Dale Carder, Univ. of Wisconsin
- NSF Program managers : Kevin Thompson and Bryan Lyles
- UVA graduate and UG students: Jie Li and Yicheng Liang
- UCAR: Mohan Ramamurthy and Pete Siemsen
- UNH: Bob Russell, co-author on VCMTP & DYNES collaborator
- Internet2: Brian Cashman & Eric Boyd
- V. Tech: John Lawson and Jeff Crowder; UVA: Ron Withers
- MAX: Tom Lehman
- IU: Ezra Kissel and Martin Swany
- RENCI: Chris Heermann; Texas Tech: Alan Sill; CSU: Greg Redder
- BU: Weylin Piegorsch & David Starobinski
- U. Utah (Emulab), UNM PRObE, and ESnet 100G testbed support

Wide-area distributed instrument: DYNES

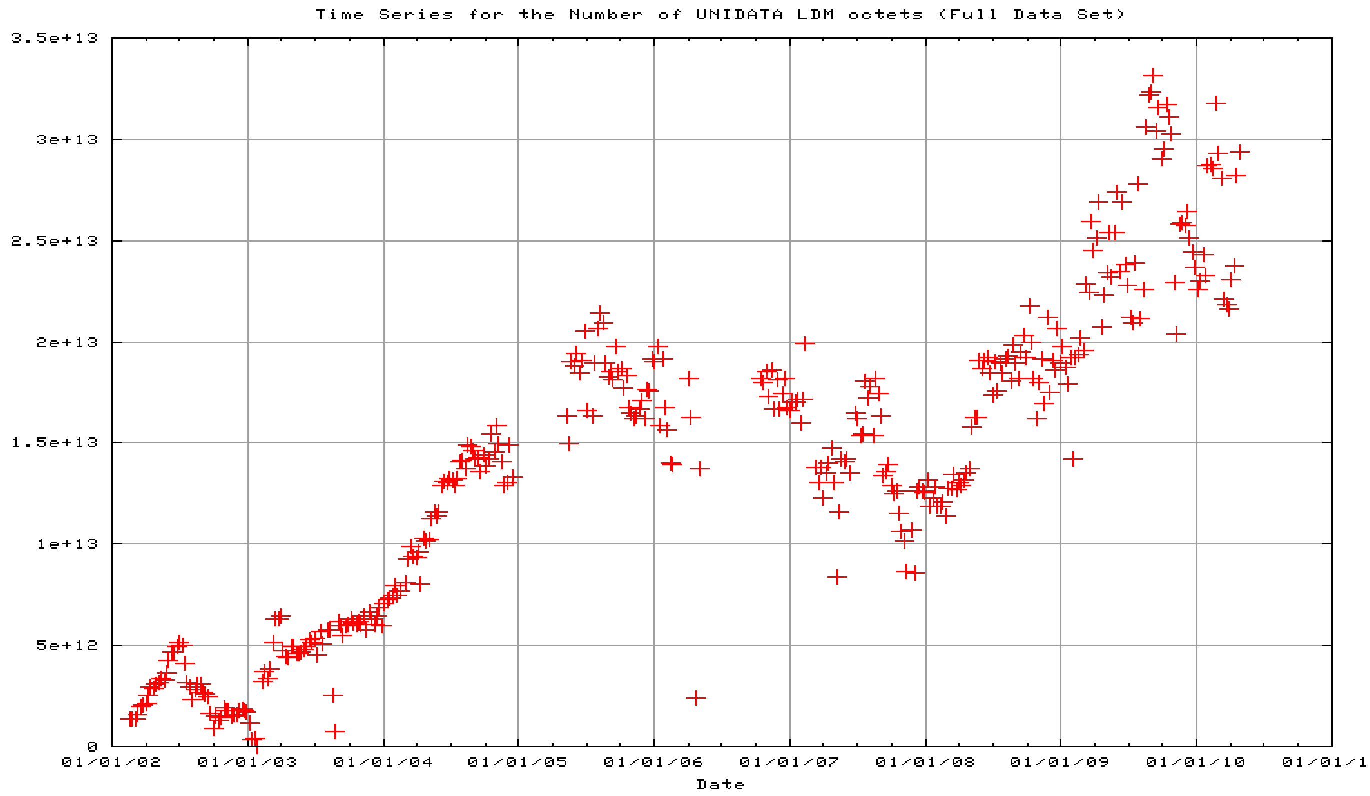


- Internet2 AL2S and ION services
- Regional services: DYNES or stitched VLANs, e.g., MARIA, FRGP

Local Data Manager (LDM)

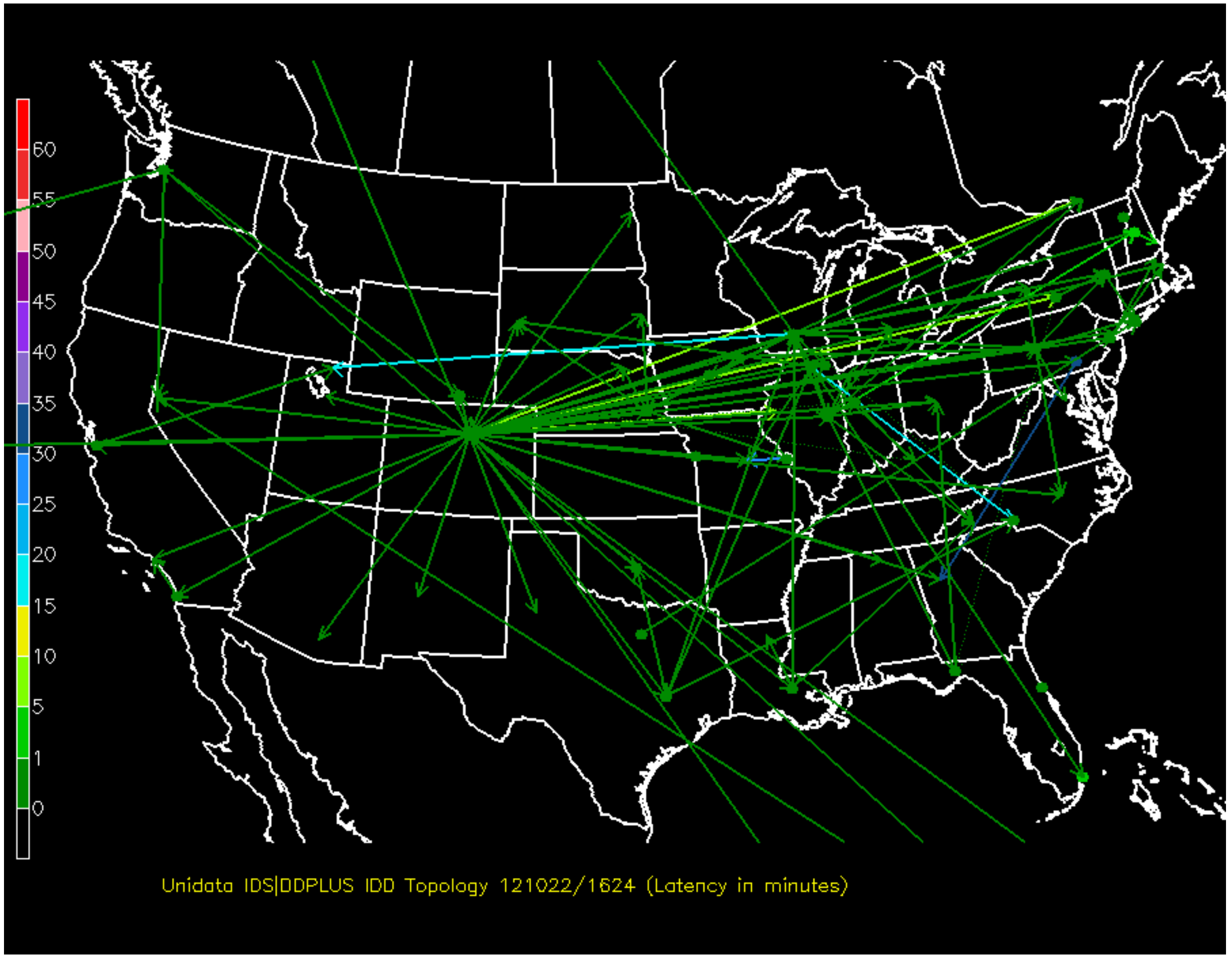
- Software used by the Unidata Internet Data Distribution (IDD) system to deliver meteorology-related data to at least 450 computers at 240 sites (mostly universities).
- "Push" (not "pull") of routinely-generated data-products
- 30 data feeds provide radar, satellite, text bulletins, lightning, model forecasts, surface and upper air observations, …
- Receives over 14 GB/h, 24/7 (usage is increasing).
- Transmits over 650 GB/h, 24/7, with 99.999% reliability (usage is increasing).
- Uses RPC over unicast TCP connections
- The National Weather Service uses LDM-6 to collect and relay NEXRAD level 2 radar data operationally for over 162 radars
- LDM-6 was recently selected for data collection for the THORPEX Interactive Grand Global Ensemble (TIGGE)

LDM Use on Internet2



- The above time-series plot shows the weekly flow of LDM traffic on Internet2 from early 2002 to 20100426.

IDD in the USA



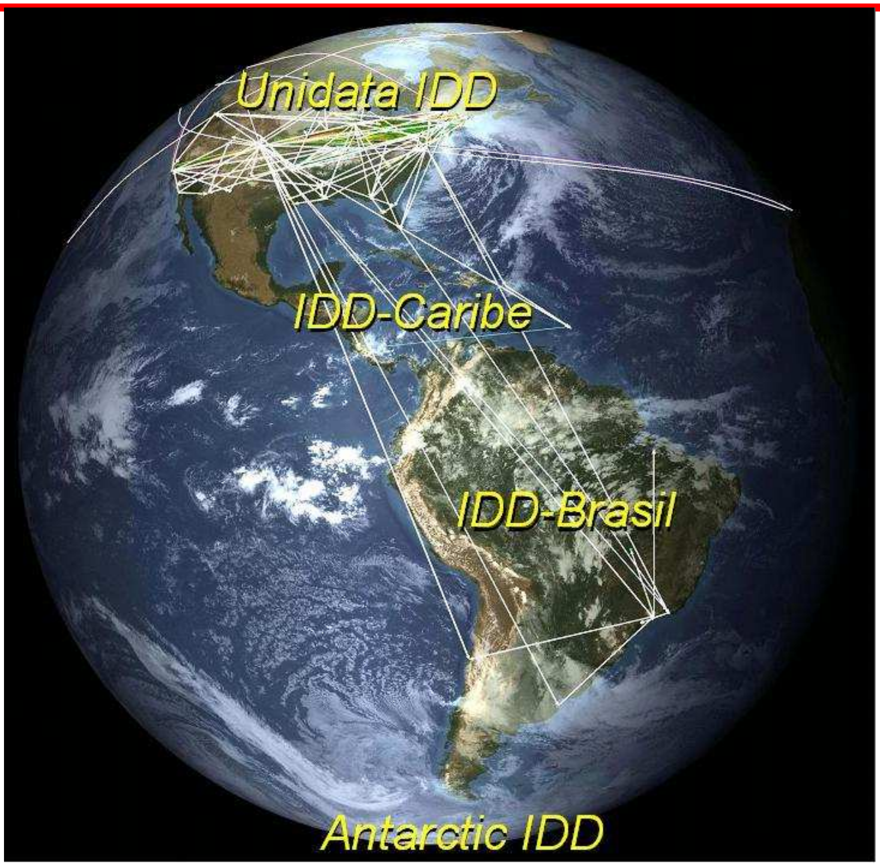
Global Internet Data Distribution

Unidata IDD
North American data delivery and sharing network

IDD-Brasil
South American peer of North American IDD

IDD-Caribe (planning)
Central American peer of North American IDD

Antarctic-IDD
Support of US Antarctic research community



Participants

United States
Canada
Puerto Rico
Costa Rica
Barbados
Venezuela
Chile
Brazil
Argentina
England
Portugal
Spain
Austria
Russia
Vietnam
China (Hong Kong)
South Korea
Antarctica (incipient)

Layer-2 vs. IP vs. AL multicast

- Compared to AL multicast
  - network switches/routers making copies should (ideally) result in lower CPU and bandwidth resources for data distribution (for same latency)
- L2 virtual circuits vs IP datagrams
  - Problems with IP multicast
    - complexity of IP multicast routing
    - receivers without credentials joining
    - congestion related packet loss
  - L2 virtual circuits have setup phase and rate guarantees
- Issues with L2 paths
  - lack of ubiquitous deployment (DYNES)
  - VC setup delay (but our app. has continuous traffic)

Virtual Circuit Multicast Transport Protocol (VCMTP)

- VCMTP was designed for the type of data distributed by LDM
  - Almost continuous streams of small files
  - Hundreds of receivers not millions
- VCMTP design choices
  - Reliable multicast over Layer-2 virtual circuits
  - Retransmission requests/retx over TCP
  - Tradeoff throughput of fast receivers with robustness of slow receivers
    - Knob: per-file retransmission timeout factor
  - Asynchronous and message-based API
  - Receiver: per-file notification or batched notification

Developed a reliable multicast transport protocol & tested on U. Utah Emulab (500 node cluster)

TABLE 7: Experiment 2 results (continuous file transfers)

	ρ	Loss Rate	Timeout Factor	Robustness R as a percentage (SD)			Throughput T in Mbps (SD)		
				$n = 10$	$n = 20$	$n = 30$	$n = 10$	$n = 20$	$n = 30$
Config. 1	0.4	5%	10	86.3 (2.3)	83.4 (0.4)	81.4 (0.7)	92.8 (1.1)	90.5 (0.7)	86.9 (0.7)
Config. 2	0.4	5%	50	98.9 (0.8)	98.1 (0.7)	97.5 (0.3)	92.7 (0.8)	89.2 (0.7)	85.8 (0.8)
Config. 3	0.4	10%	10	79.9 (2.3)	74.0 (1.2)	65.2 (0.5)	90.5 (0.8)	85.1 (0.6)	82.3 (1.0)
Config. 4	0.4	10%	50	96.2 (1.9)	94.0 (1.6)	85.0 (1.3)	89.9 (0.5)	84.9 (0.7)	80.3 (1.6)
Config. 5	0.8	5%	10	35.0 (4.7)	25.9 (3.5)	15.7 (3.8)	93.9 (1.0)	92.0 (0.5)	91.7 (0.8)
Config. 6	0.8	5%	50	68.2 (6.0)	60.3 (5.5)	55.8 (6.0)	92.1 (0.8)	88.9 (2.7)	88.4 (1.5)
Config. 7	0.8	10%	10	22.9 (4.4)	11.6 (3.2)	10.6 (1.5)	93.6 (0.3)	91.1 (0.6)	89.2 (1.2)
Config. 8	0.8	10%	50	56.3 (9.5)	53.3 (1.7)	50.2 (3.6)	92.7 (1.5)	88.1 (1.8)	85.0 (1.6)

- Data multicast has to be reliable; so need unicast connections just for retransmissions
- Timeout factor: retx requests received after timeout-factor*file-multicast time will be rejected
- robustness: ratio of successfully received files at lossy receivers
- throughput: computed for lossless receivers

Key findings

- Robustness increases significantly when timeout factor is increased without a significant drop in throughput when traffic load is low
- At high traffic loads, increasing timeout factor from 10 to 50 only increased robustness to 50%; lower load by
  - adding more sender capacity, or
  - moving lossy (multitasking) receivers to lower rate virtual circuit

Current work/plan

- Integrating VCMTP with LDM
- Test on UNM PRObE with 200 receivers
- Develop RoCE version and test on ESnet 100G testbed: Bob Russell, UNH
- Compare AL, IP and Layer-2 multicast across DYNES hosts (ten planned)

Questions/comments?

email: mvee@virginia.edu