Getting it Together: Enabling Multi-organizational Provenance Exchange

M. David Allen, Adriane Chapman, Barbara Blaustein, Len Seligman
{dmallen, achapman, bblaustein, seligman}@mitre.org

MITRE Sponsored Research
The Story

Capture Agent

Coalition Provenance Store

NGO Provenance Store

Marine Corps Provenance Store

Federated Provenance Manager

REPORT

STORE

PUBLISH

RETRIEVE

Official Thai Armed Forces photo by Thai Army Master Sgt. 1st Class Nuttrawut Thopradid/released; www.marines.mil
Desiderata for Provenance Exchange

- **Scalability**
  - Very high limit (or none) to the number of participants

- **Agility**
  - When new participants join the network, other participants are not required to change software or have knowledge of the new participants

- **Autonomy**
  - Providers retain tight control over their own data, making it easier for them to decide to share

- **Resiliency**
  - Even when some stores fail or leave the sharing community, the overall sharing capability remains intact
Not Your Standard Data Exchange

Problem

- Inability to specify all nodes and edges at initial query
- Only able to specify edges incident to current node
- Must spider out, generating queries on the fly
- No store owner has the whole result
How to Accomplish it

Assembling the complete picture requires cooperation

Shipboard Provenance Store

Air Force Provenance Store

Joint Operations Provenance Store

Provenance of 5

Provenance of 4

Response:

1

2

3

4

5

Response:

1

2

3

4

5

Overlay Network

Service Layer

Service Layer

Service Layer

Photo credits: U.S. Navy photo by Photographer's Mate Airman Sheryl Campbell/Released; U.S. Navy photo by Mass Communication Specialist 2nd Class Mark Logico/Released; U.S. Marine Corps photo by Master Sgt. Grady T. Fontana/Released)
Peer-to-Peer Architecture

- **Scalability, Agility**
  - Each new participant agrees to expose standard services
  - Standard P2P routing techniques support discovery over all other peers in the network

- **Autonomy**
  - Each individual system only exposes information to the participants it authorizes

- **Resiliency**
  - No single coordinating authority
Initial Performance Experiments

Number of Provenance Objects Returned vs. time (ms)

Discovery

Query and Retrieve

# advertisements related to object vs. time (ms)