

Visual Representations of Flow Data

and the Value of Visual Language

Presented by Sunny Fugate
Space and Naval Warfare Systems Center, San Diego



Human-Machine Efficiency

Over-Learned: **Feedback**

non-volitional feedback



haptic

volitional feedback



visual / aural

Human-Machine Efficiency

Over-Learned: **Feedback**

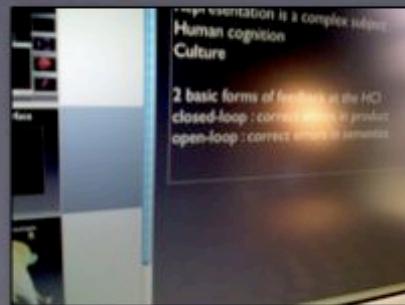
non-volitional feedback



correct errors in production

haptic

volitional feedback



visual / aural

Human-Machine Efficiency

Over-Learned: **Feedback**

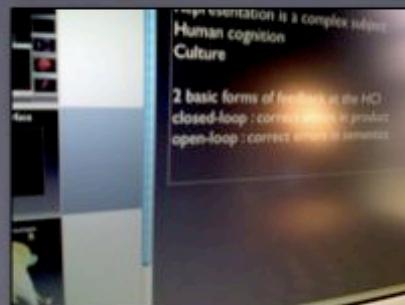
non-volitional feedback



correct errors in **production**

haptic

volitional feedback



correct errors in **semantics**

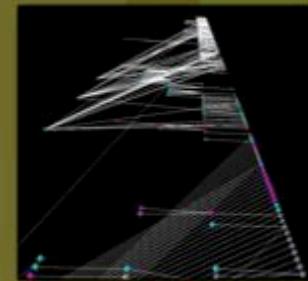
visual / aural

Human-Machine Efficiency

Over-Learned: **Feedback** - haptic vs visual/aural

Haptic Feedback

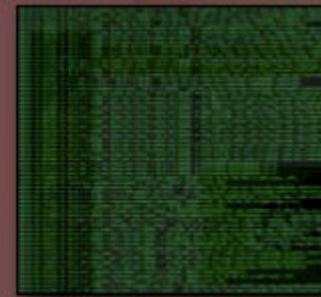
Sequential access



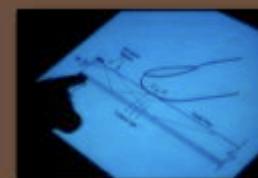
Random access



keyboard



data/gesture glove



voice control

Human-Machine Efficiency

Over-Learned: **Feedback** - haptic vs visual/aural

Haptic Feedback

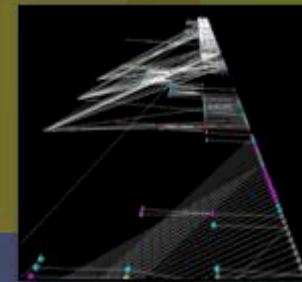
Sequential access



Sidewinder™
Force Feedback



Falcon™



joystick



mouse

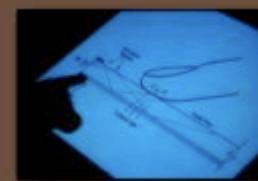
Random access



keyboard



data/gesture glove



multi-touch



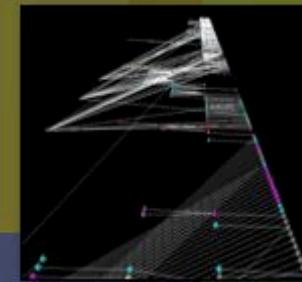
voice control

Human-Machine Efficiency

Over-Learned: **Feedback** - haptic vs visual/aural

Haptic Feedback

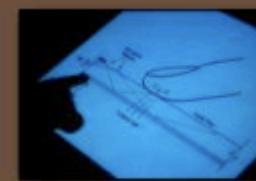
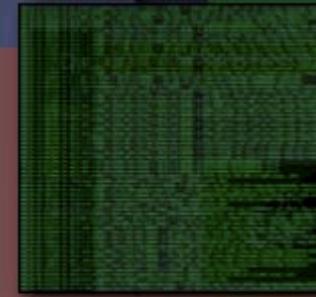
Sequential access



Random access



keyboard



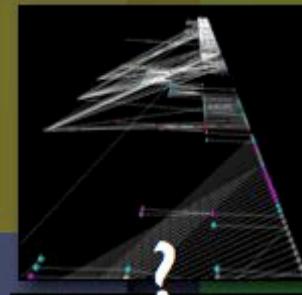
voice control

Human-Machine Efficiency

Over-Learned: **Feedback** - haptic vs visual/aural

Haptic Feedback

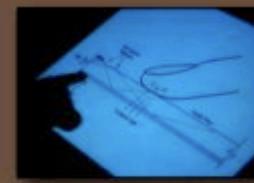
Sequential access



Random access



keyboard



multi-touch



voice control

Human-Machine Efficiency

Under-Learned: Representation

arbitrary



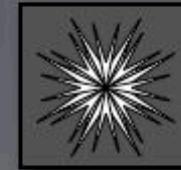
PCAP

TXH 1138

association



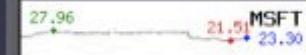
metaphor



representational



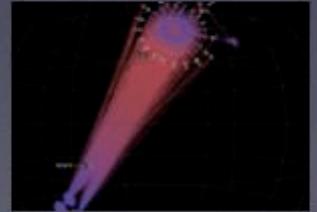
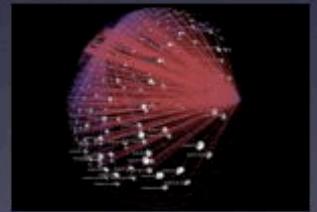
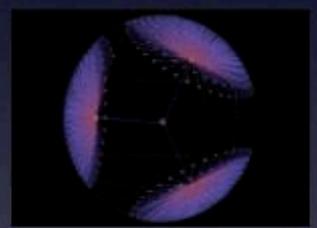
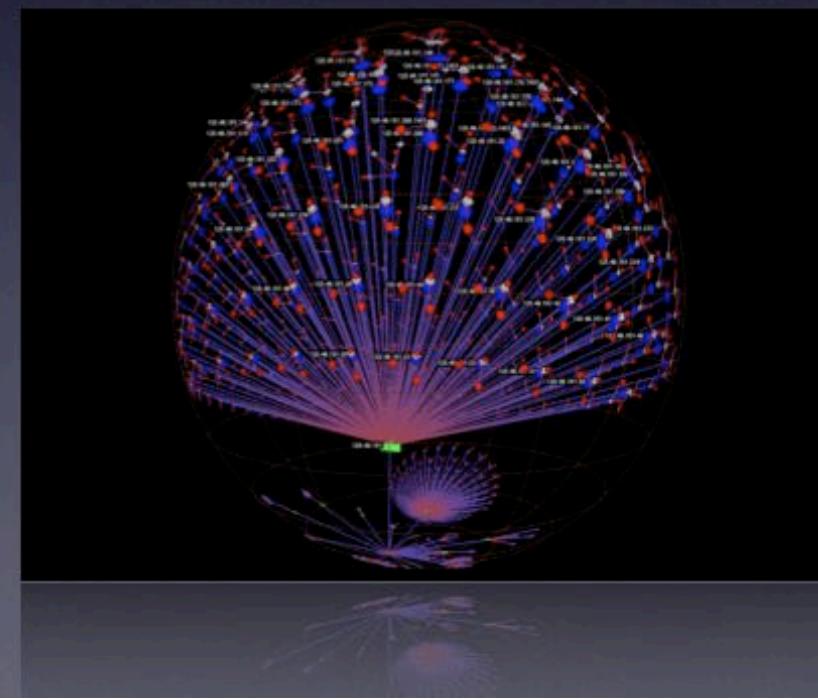
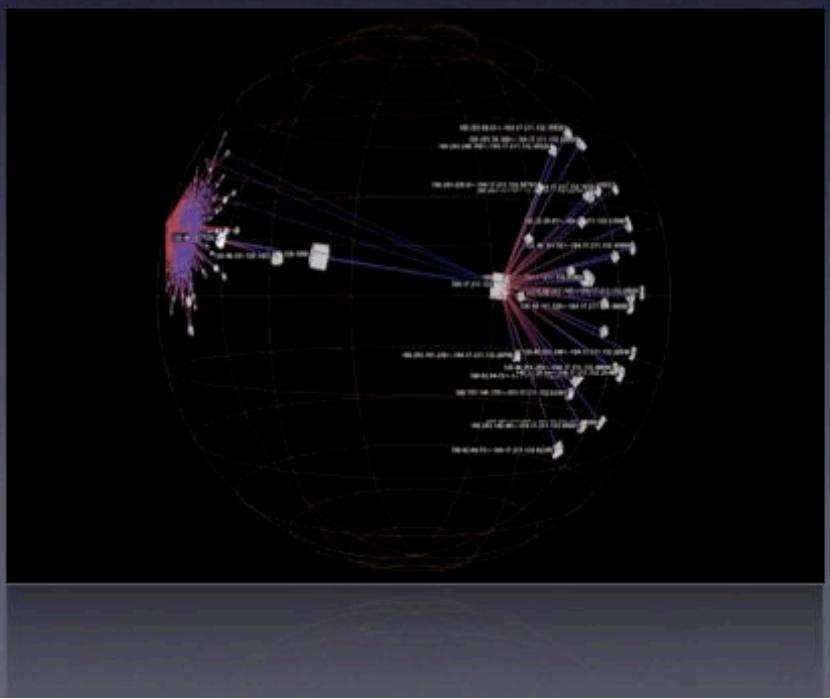
indexical



Culture/Domain Specificity

Flow in hyperbolic space

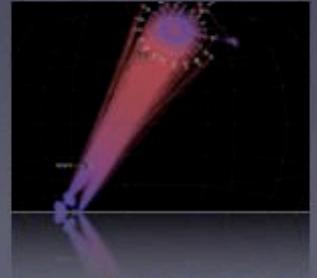
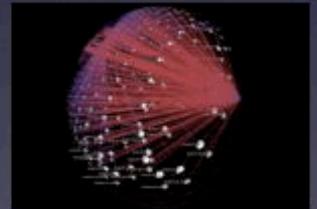
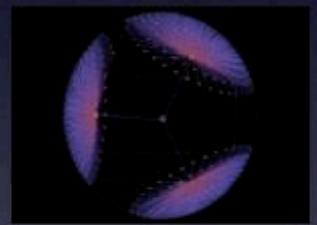
- 3 month SSC project in 2002
- discover and apply network visualization tools
- **Hyperviewer**: quasi-hierarchical hyperbolic space
- **'fish-eye'** 3-d
- Created by Stanford researcher **Tamara Munzner**

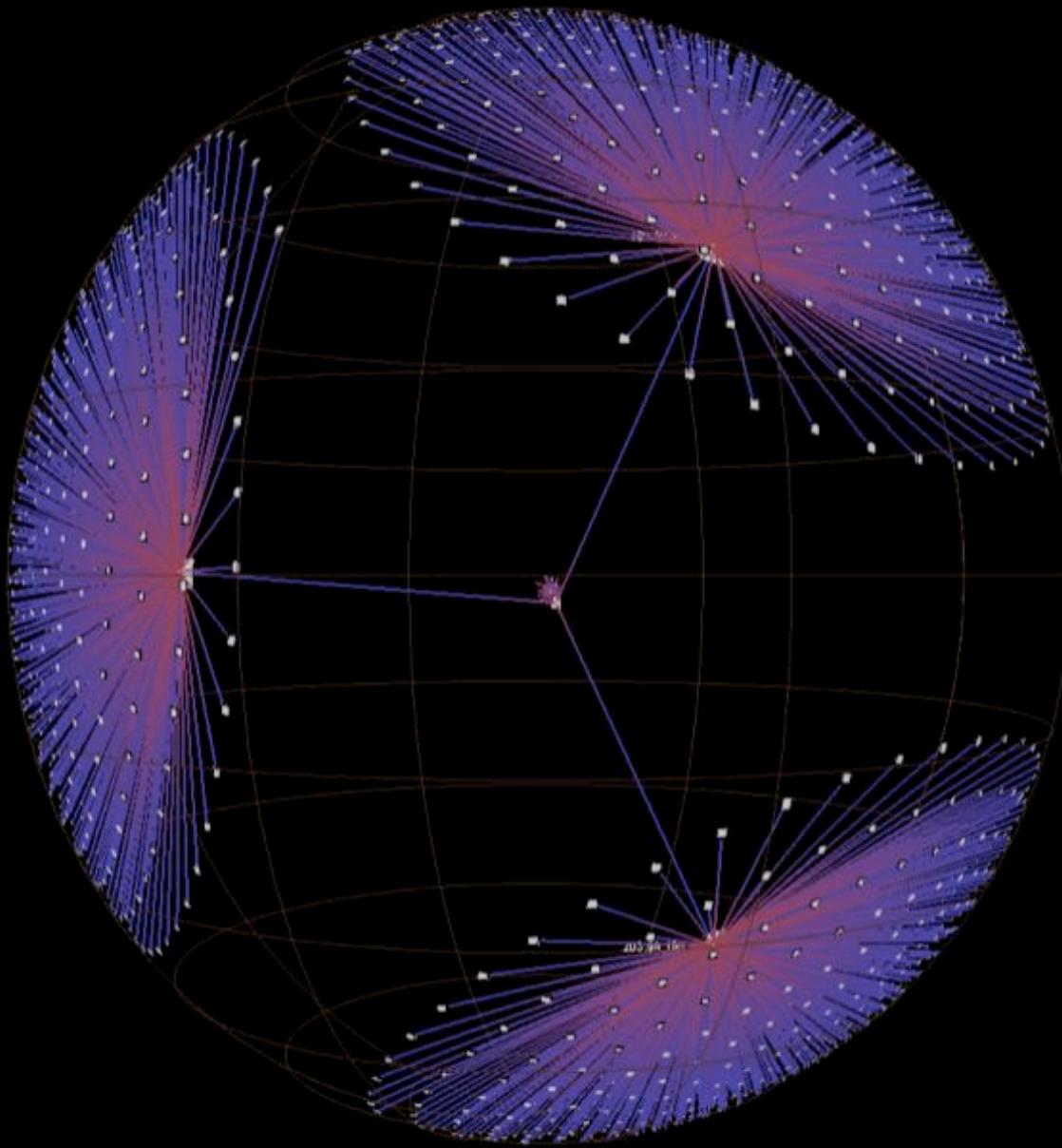


Flow in hyperbolic space

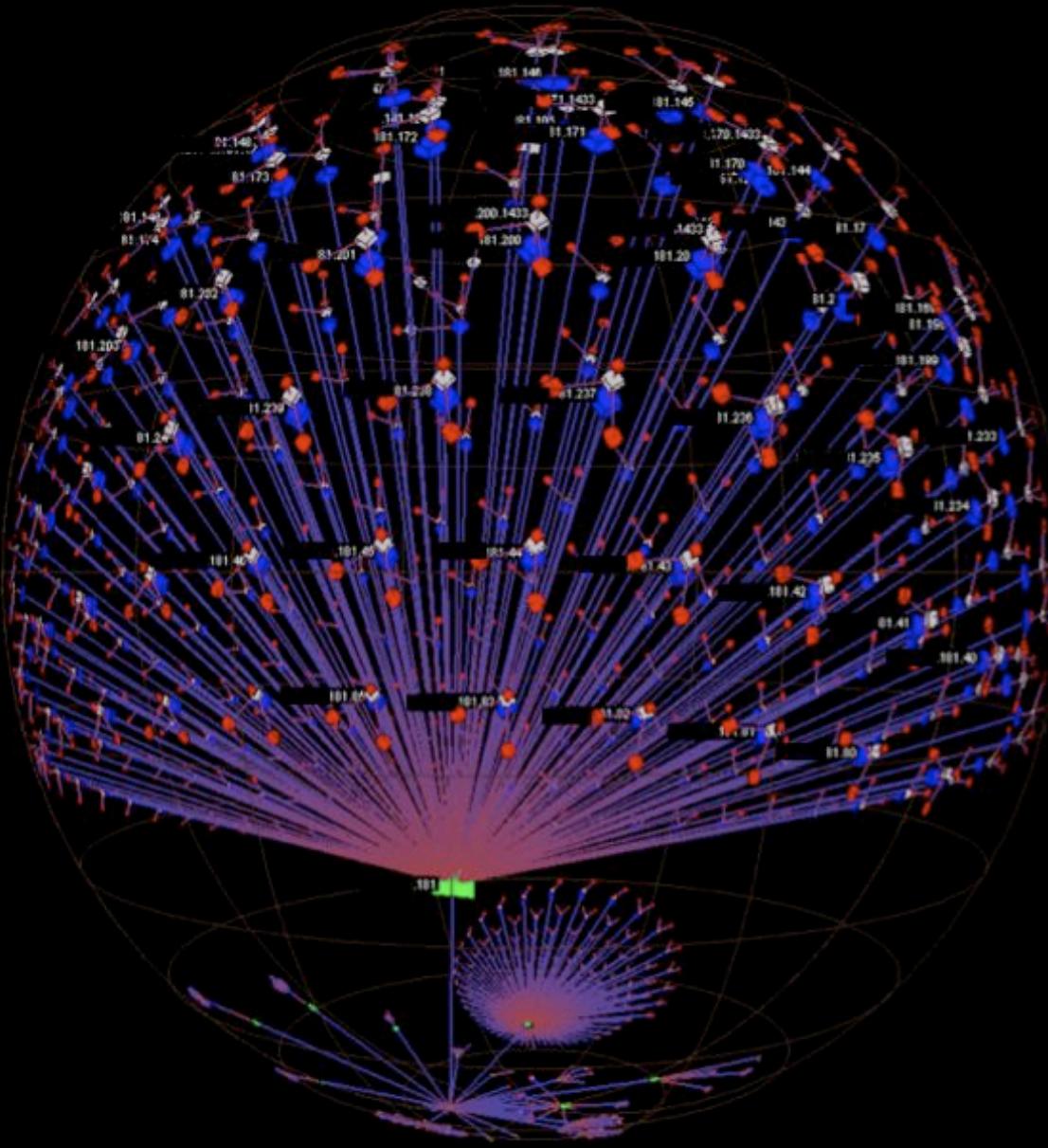
- Easily adapted to a forced-hierarchy view of flow
- **Opensource C++ library and UI**
- Experimented with visual methods

- colors
- graph cycles
- scaling
- text labels
- **graph size**
- search automation

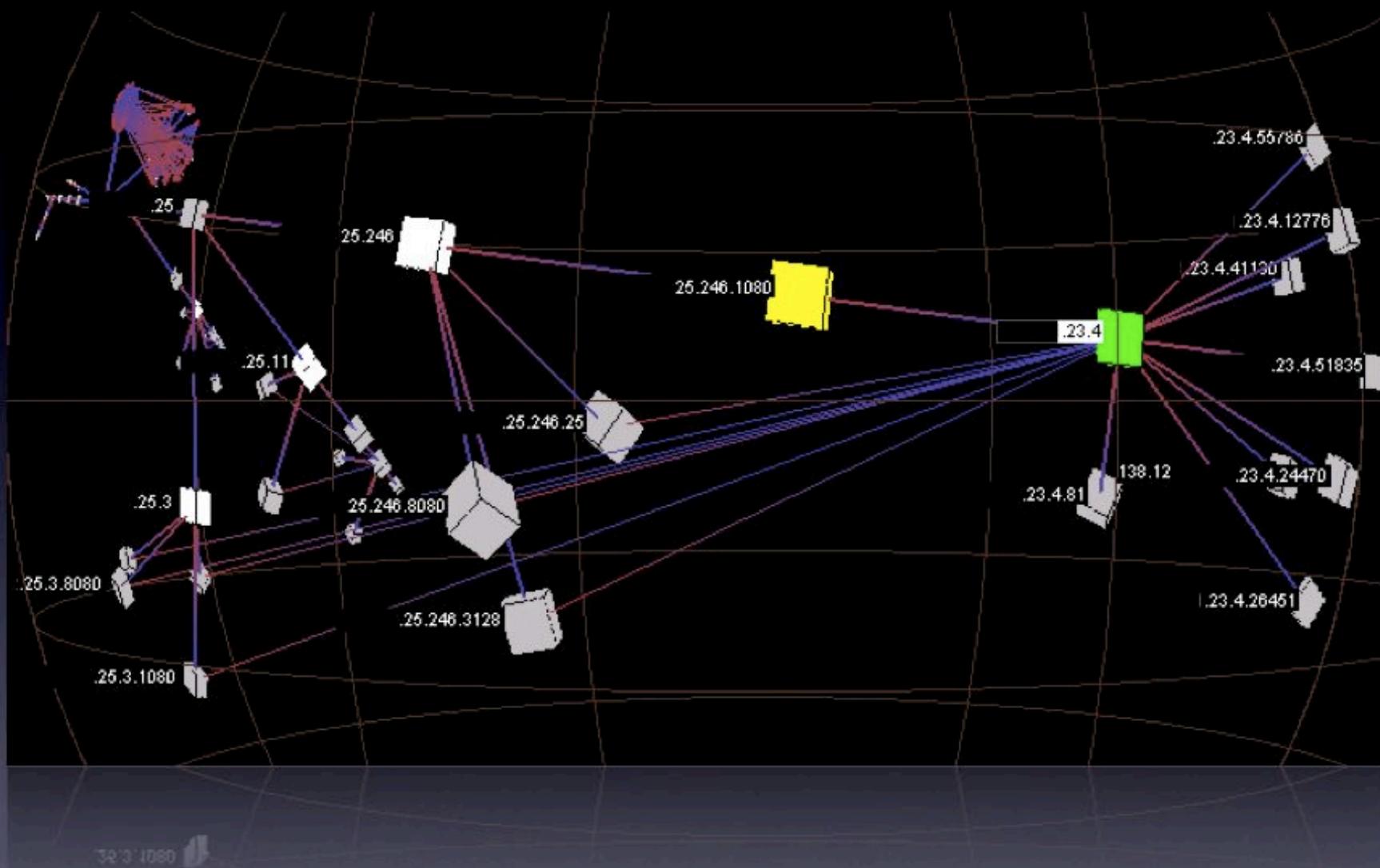




Symmetry in port access from 3 separate clients.



src/dst ports colored red/blue



Hierarchy showing client subnet and server ports

Shapes Vector

- Acquired by DARPA in 2002
- Developed by Australian DSTO
(Defence Science Technology Organisation)
- JTF-GNO pilot program from 2003-2006

What is it?

- Intelligent Agents gather information and produce inferences
- Gathers information from multiple sources
 - pcap, flow, Snort, syslog, etc
- IAs performs automated data correlation & knowledge extraction
- Integrates visual and command-line analysis
- Integrated visualization makes use of human vision
- Supports visual analysis and decision-making



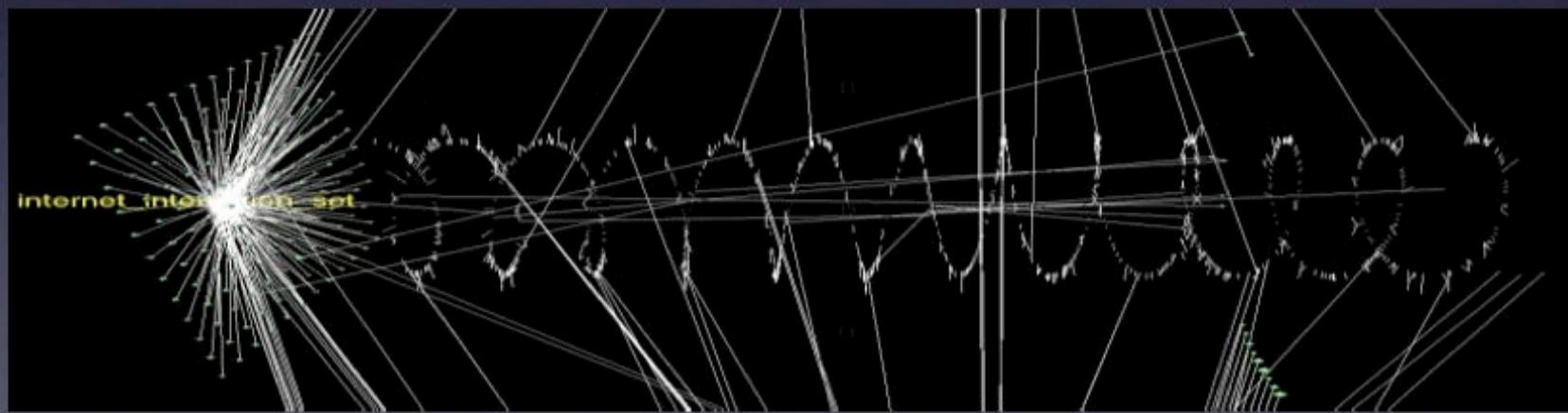
Shapes Vector

Contextual spatial, temporal, social, topological

Spatial physical geography or metaphor

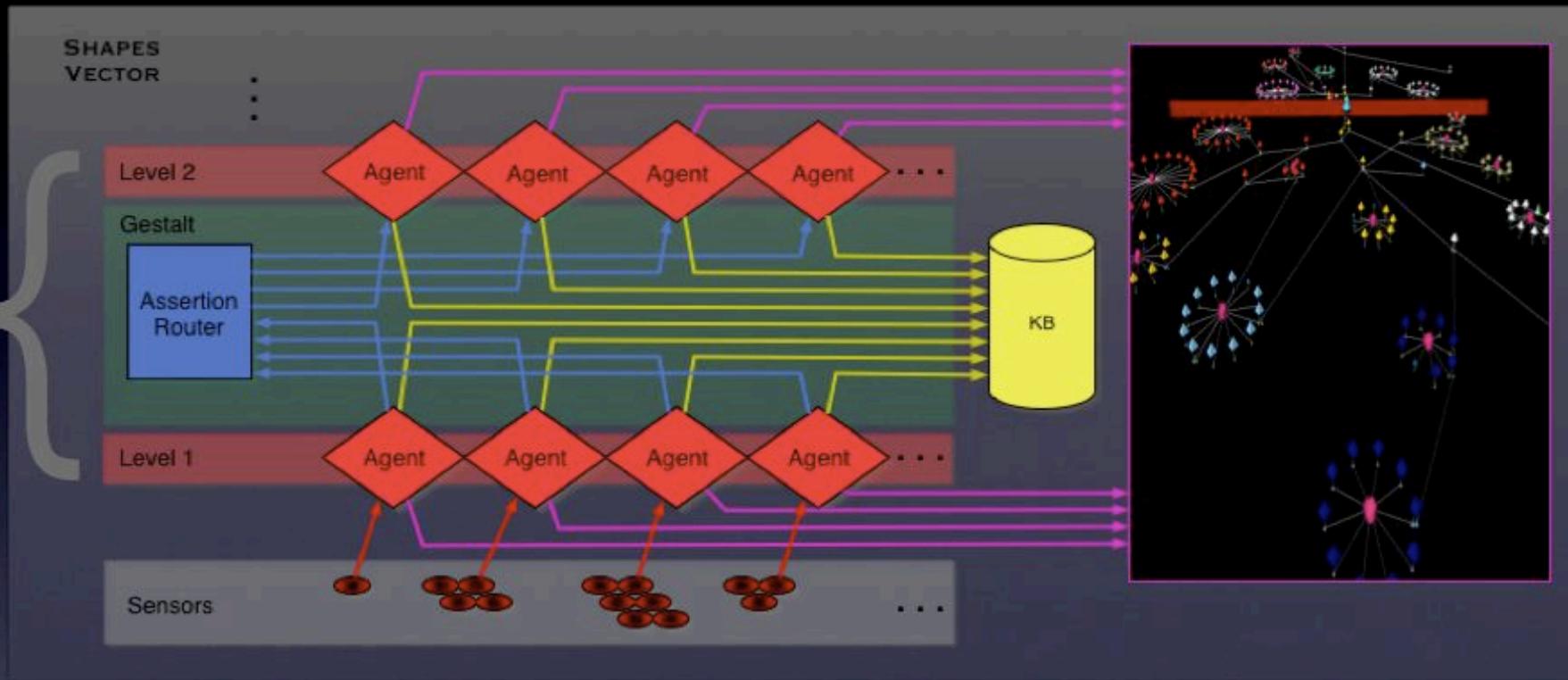
Temporal sequences in time, correlated

Visual use visual language to depict objects & events

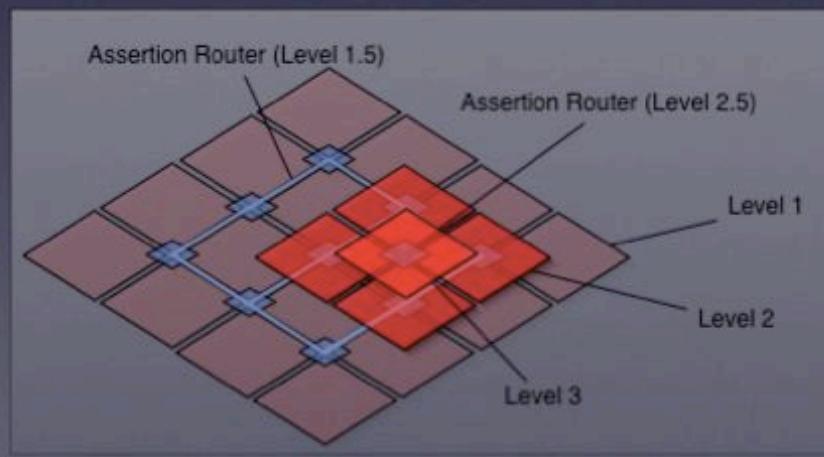


Architecture

SVKA



SVKA



- Agents can be written in many languages - must conform to the SV ontology and knowledge architecture (SVKA) specification
- Sensors can be built to wrap many information sources - must produce SV ontology
- SV ontology is a knowledge description language for network defense

Shapes Vector - Visual Language

- Easily defined visual mappings
- No applied theory of visual language

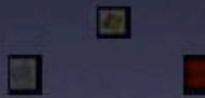
shape/color/scale



texture/icon



connection / topology



movement



packet events, information exchange, attribute changes, attribute values, host id, software, processes, machine purpose, network topology, social topology, intrusion events, event type, event priority, client vs server, routing, ...

Shapes Vector - Visual Language

- Easily defined visual mappings
- No applied theory of visual language

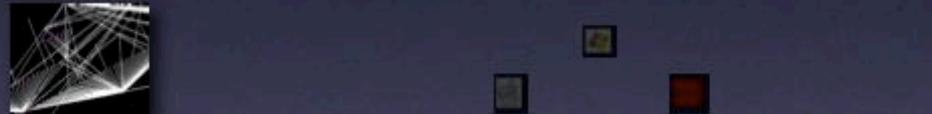
shape/color/scale



texture/icon



connection / topology



movement



packet events, information exchange, attribute changes, attribute values, host id, software, processes, machine purpose, network topology, social topology, intrusion events, event type, event priority, client vs server, routing, ...

Shapes Vector - Visual Language

- Easily defined visual mappings
- No applied theory of visual language

shape/color/scale



texture/icon



connection / topology



movement



packet events, information exchange, attribute changes, attribute values, host id, software, processes, machine purpose, network topology, social topology, intrusion events, event type, event priority, client vs server, routing, ...

Shapes Vector - Visual Language

- Easily defined visual mappings
- No applied theory of visual language

shape/color/scale



texture/icon



connection / topology



movement



packet events, information exchange, attribute changes, attribute values, host id, software, processes, machine purpose, network topology, social topology, intrusion events, event type, event priority, client vs server, routing, ...

Shapes Vector - Visual Language

- Easily defined visual mappings
- No applied theory of visual language

shape/color/scale



texture/icon



connection / topology



movement



packet events, information exchange, attribute changes, attribute values, host id, software, processes, machine purpose, network topology, social topology, intrusion events, event type, event priority, client vs server, routing, ...

Shapes Vector - Visual Language

- Easily defined visual mappings
- No applied theory of visual language

shape/color/scale



texture/icon



connection / topology



movement



packet events, information exchange, attribute changes, attribute values, host id, software, processes, machine purpose, network topology, social topology, intrusion events, event type, event priority, client vs server, routing, ...

Shapes Vector - Visual Language

- Easily defined visual mappings
- No applied theory of visual language

shape/color/scale



texture/icon



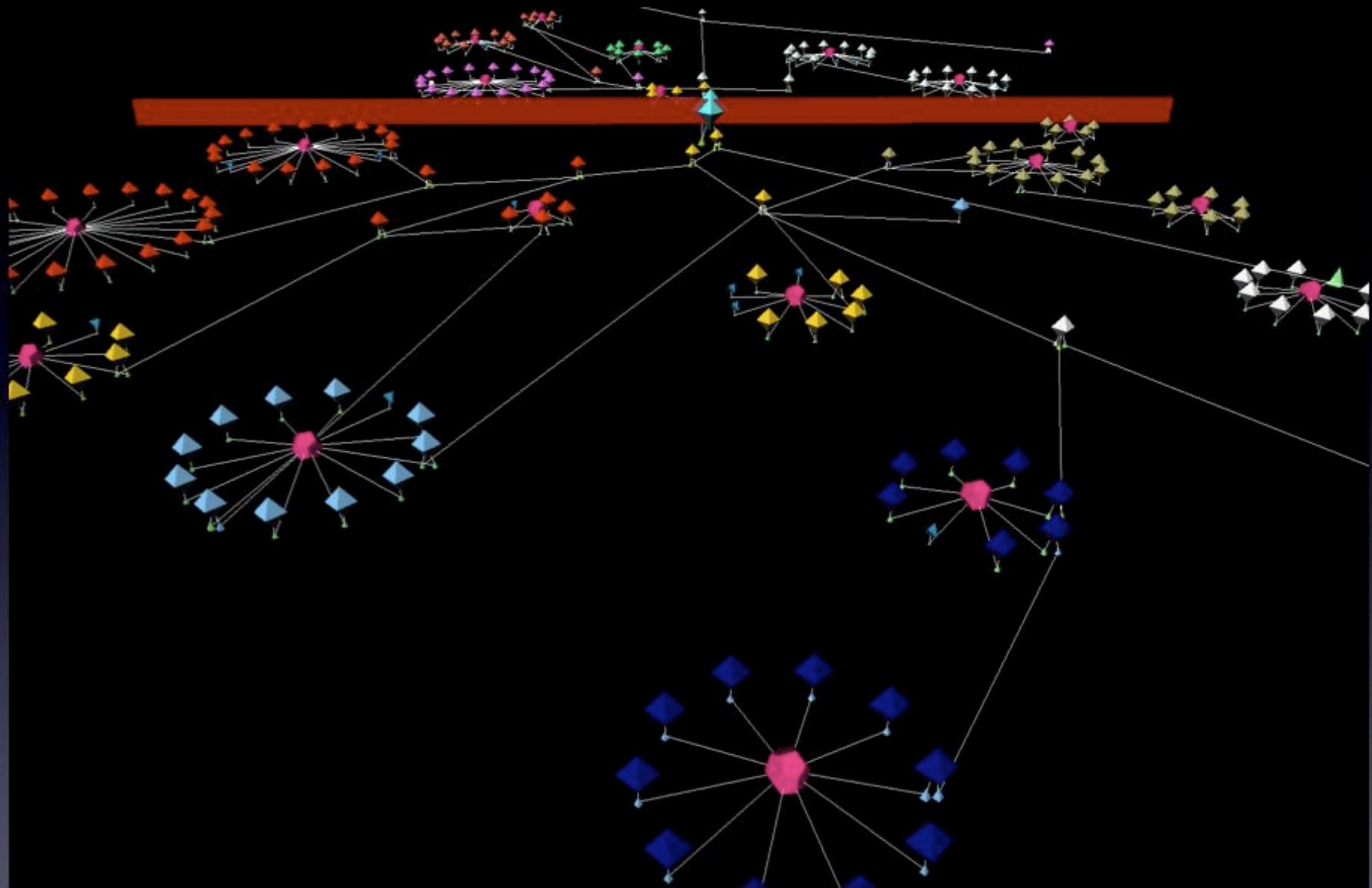
connection / topology



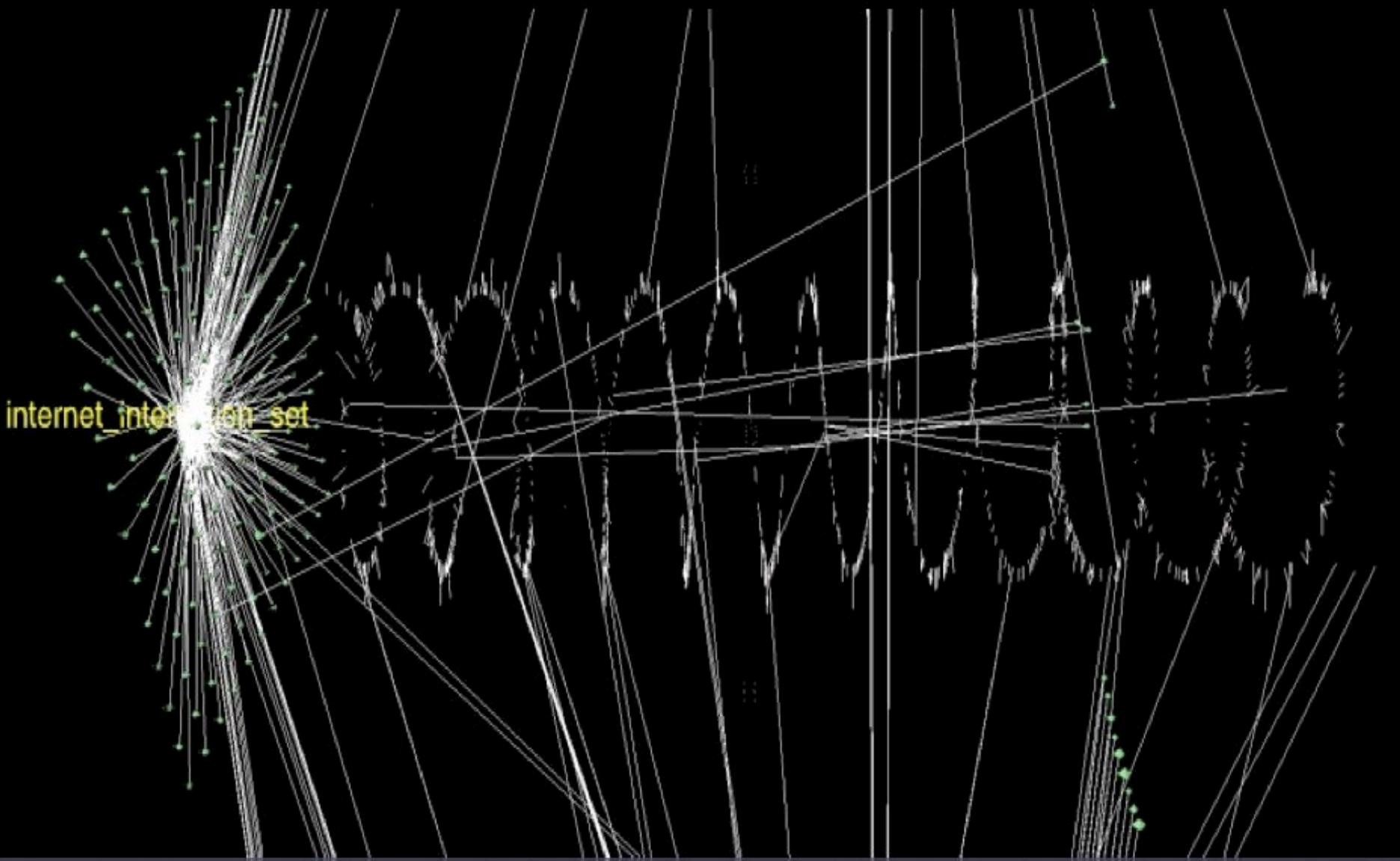
movement



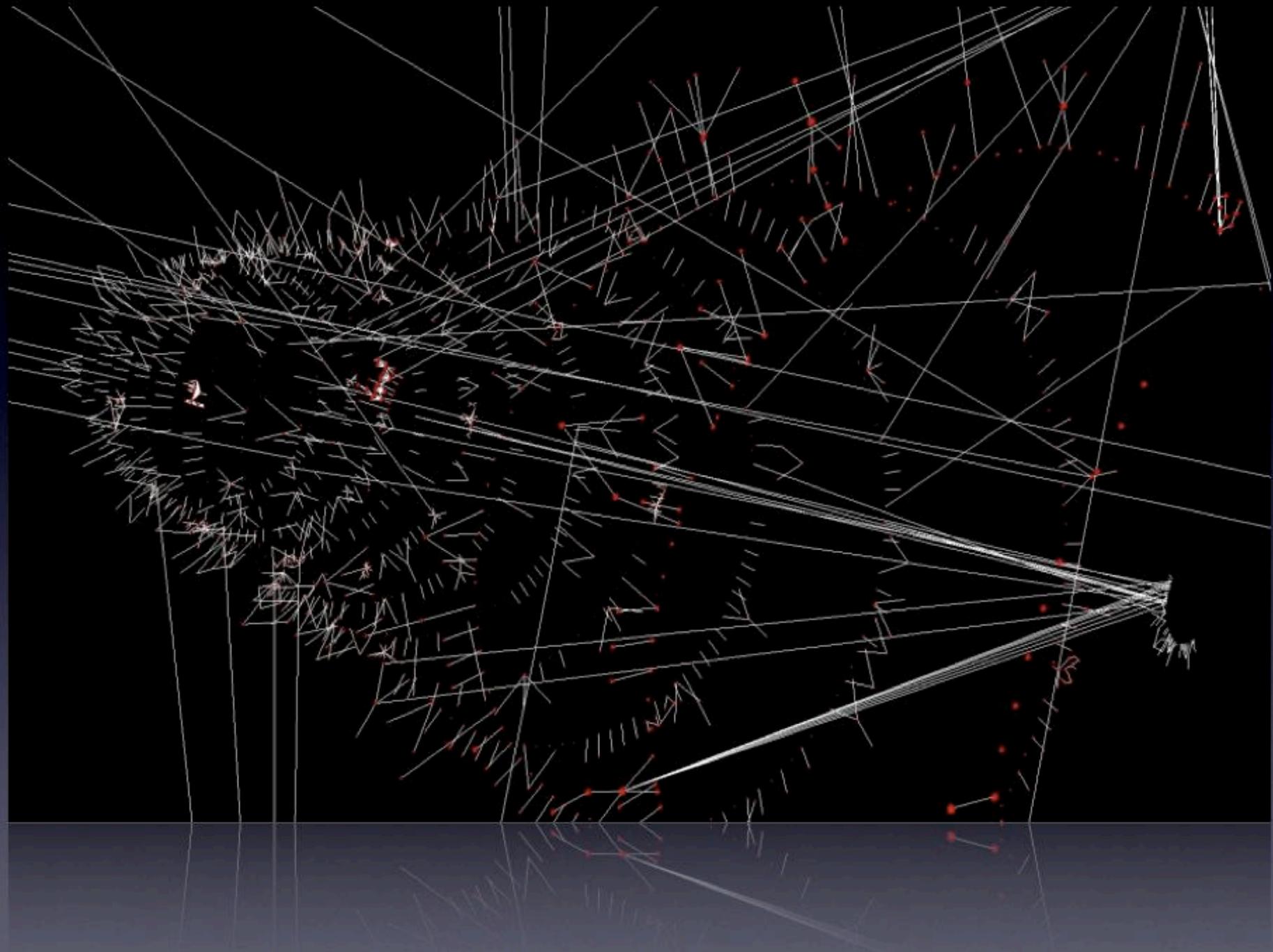
packet events, information exchange, attribute changes, attribute values, host id, software, processes, machine purpose, network topology, social topology, intrusion events, event type, event priority, client vs server, routing, ...

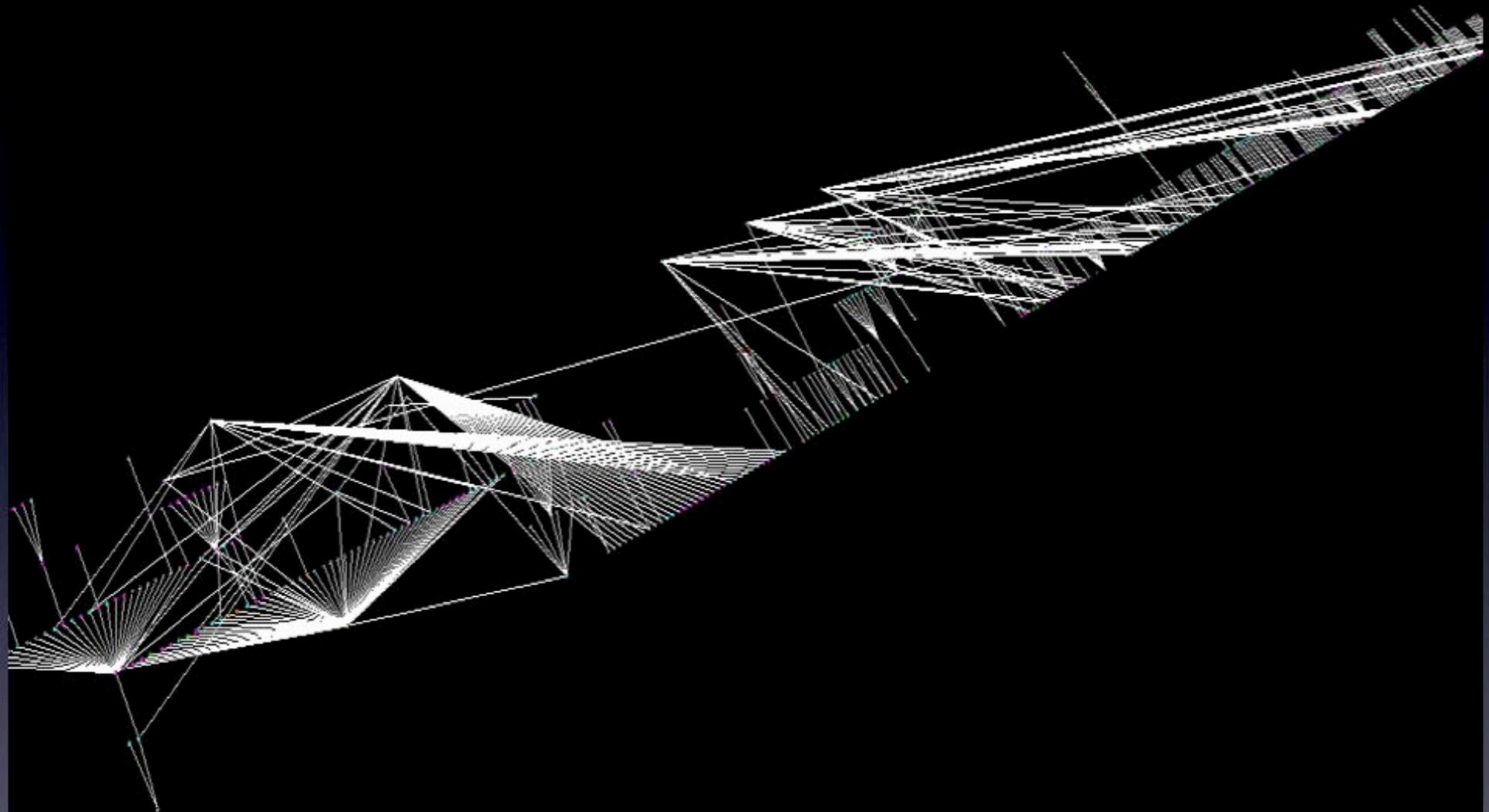


Toplogical layout using visual demarcations
(e.g. firewall, network segment, physical layout)

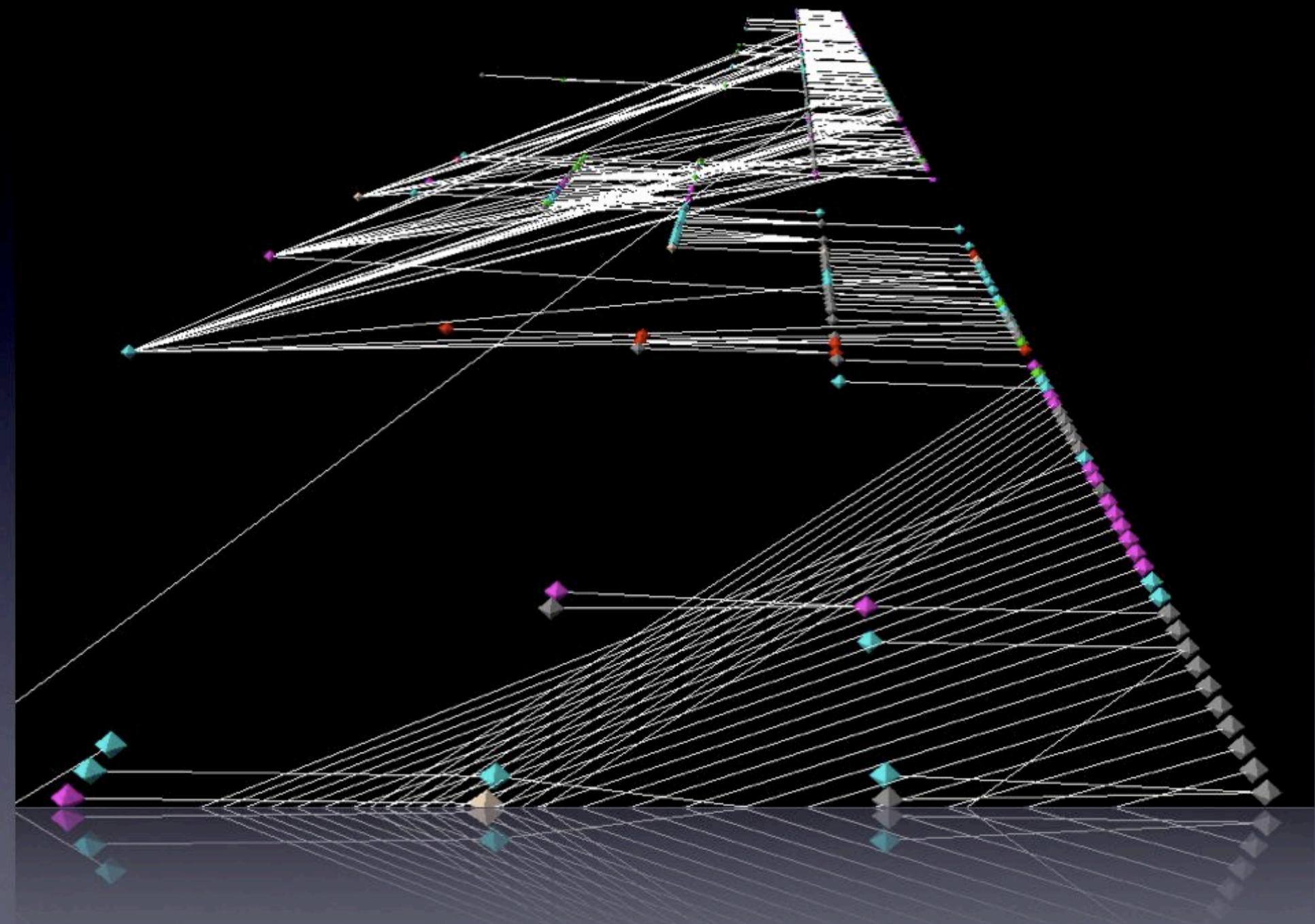


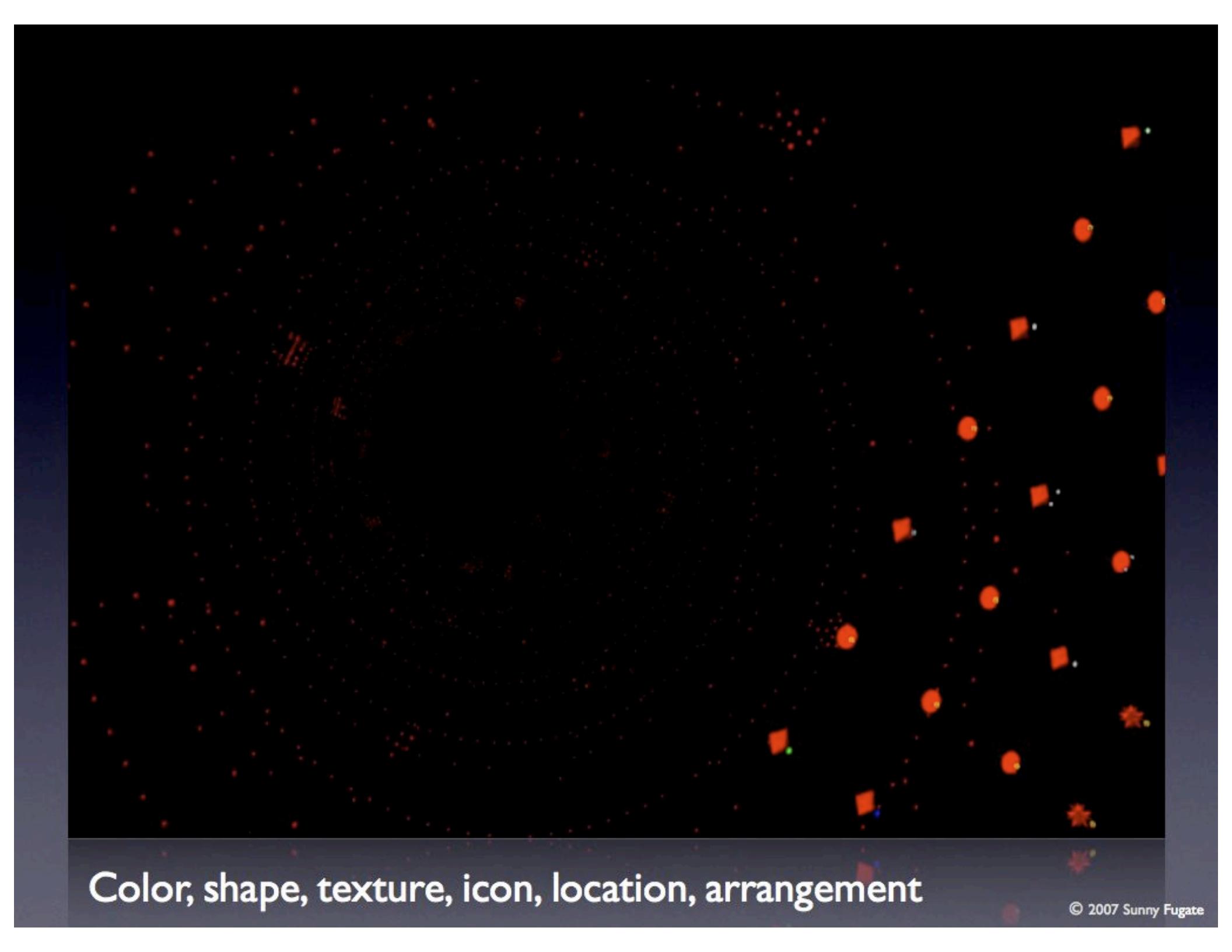
Automated layout to arrange hundreds of sub-graphs in a non-overlapping manner.



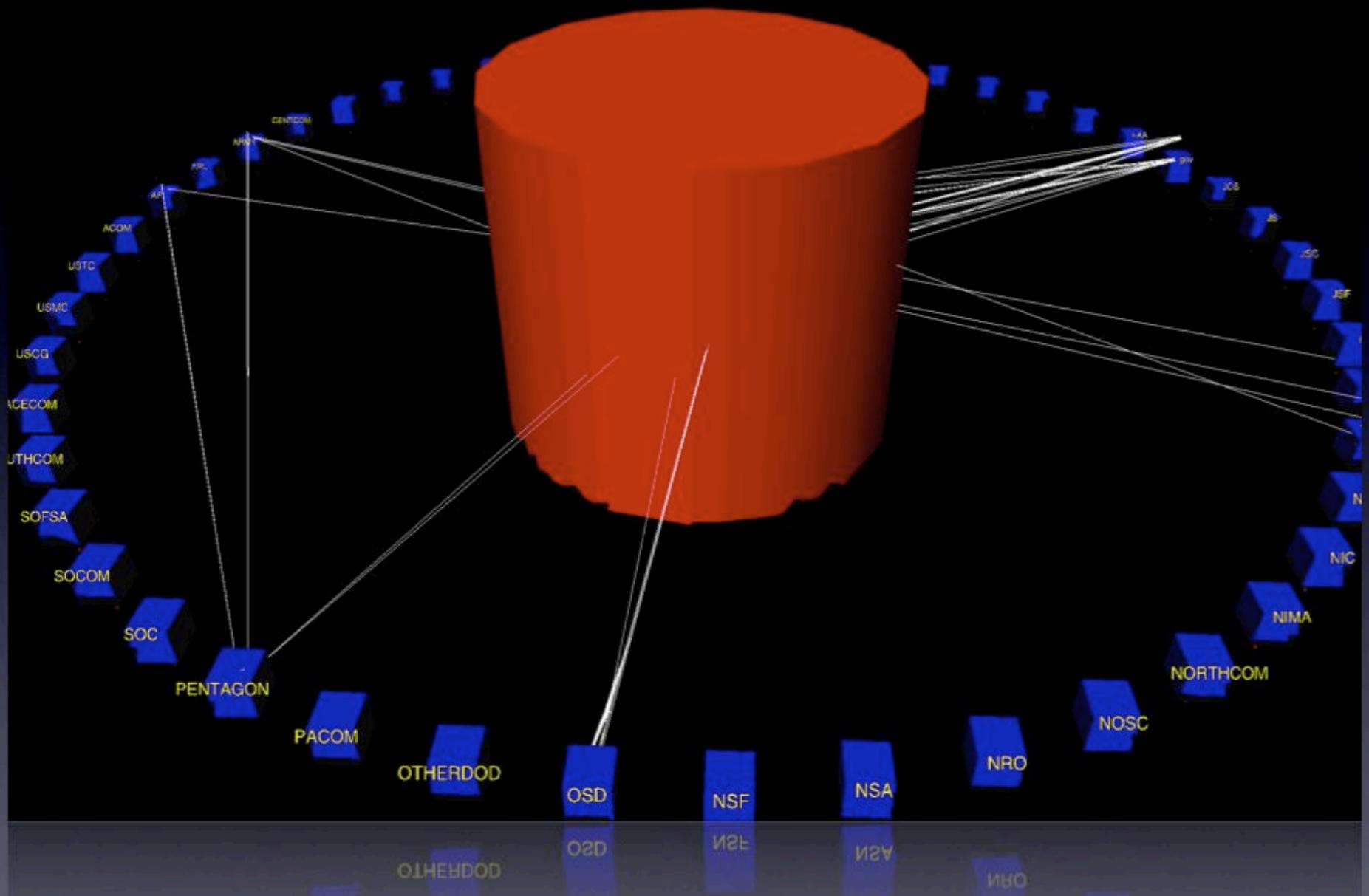


**Topological layout discovered using hints in the data
(e.g. TTL)**

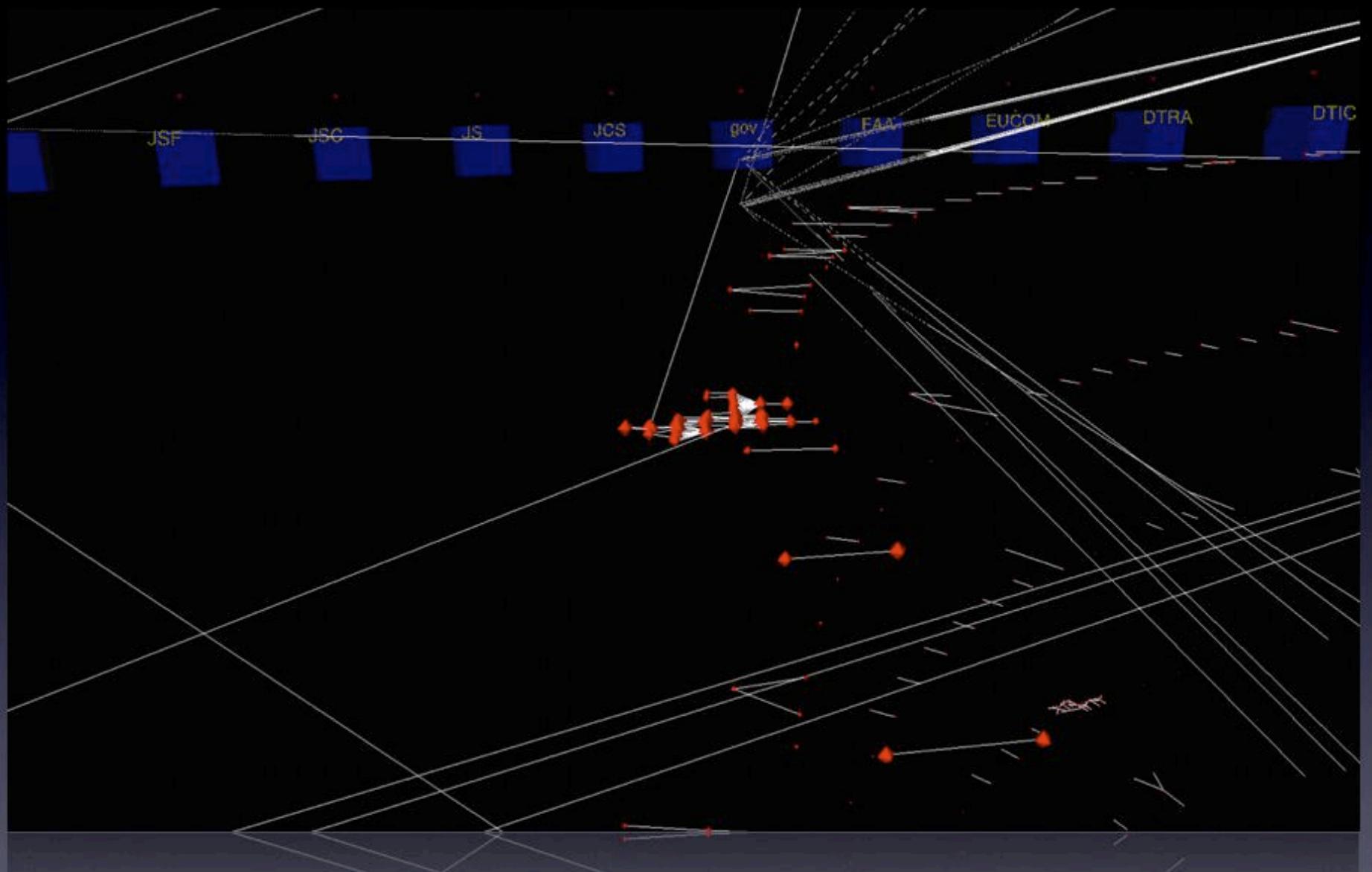




Color, shape, texture, icon, location, arrangement



Visual grouping, demarcation, and detail-hiding



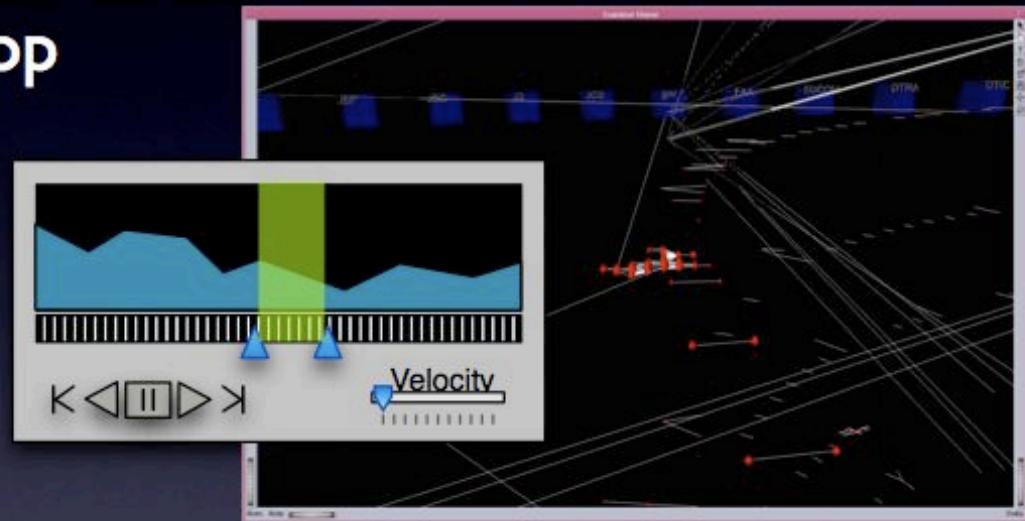
Expansive vantage points for network analysis

Shapes Vector Flow Viewer

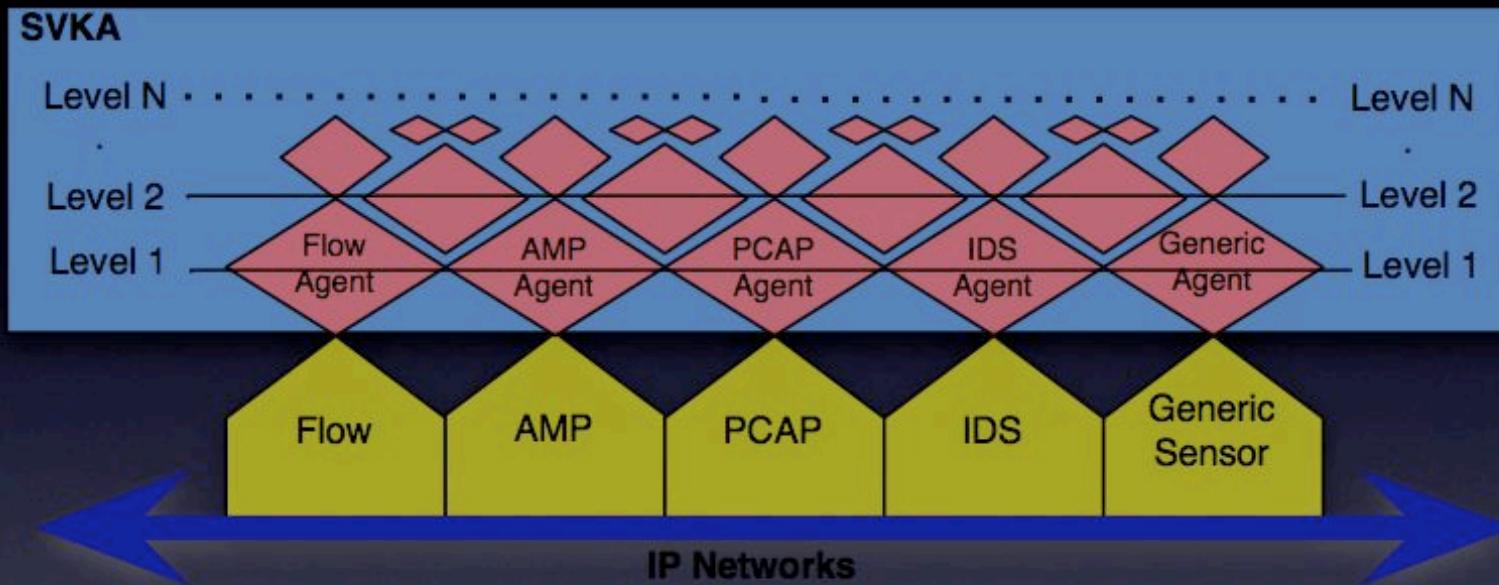
- JTF-GNO funded effort to implement SV
 - Use SV architecture and components
 - DARPA demo system > operational system
 - New scripts, sensors, agents, and GUI
- Results
 - A visual **augmentation** of CLI
 - Produces a view of **social topology**
 - **Intuitive** view of gobs of data
 - **static topology** and event **replay**
 - Links statistical views and topology view

Flow Viewer GUI

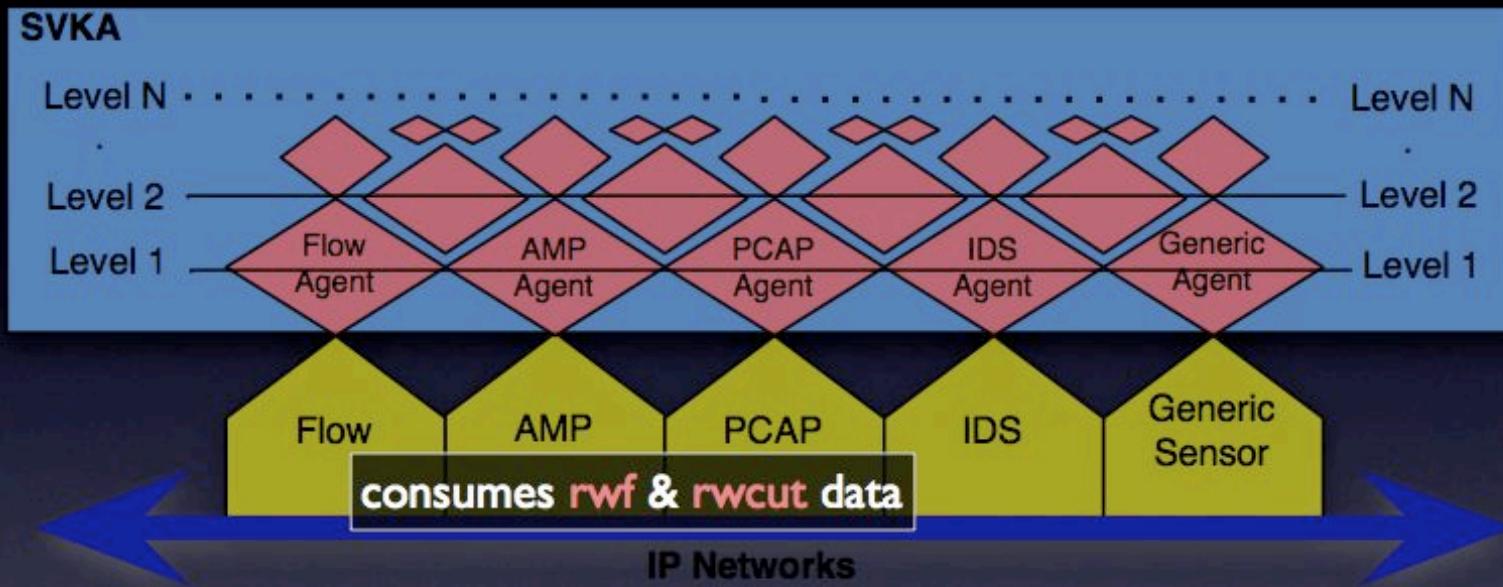
- multiple stats views linked to visuals
- playback specific ranges & loop
- adjust replay **velocity**
- time-skip
- IP and attribute **hotlists**
- dynamic **filtering** controls
 - GUI managed **rwfilter**
 - filter using SV **ontology**
- integration between **flow, AMP, IDS, & PCAP**



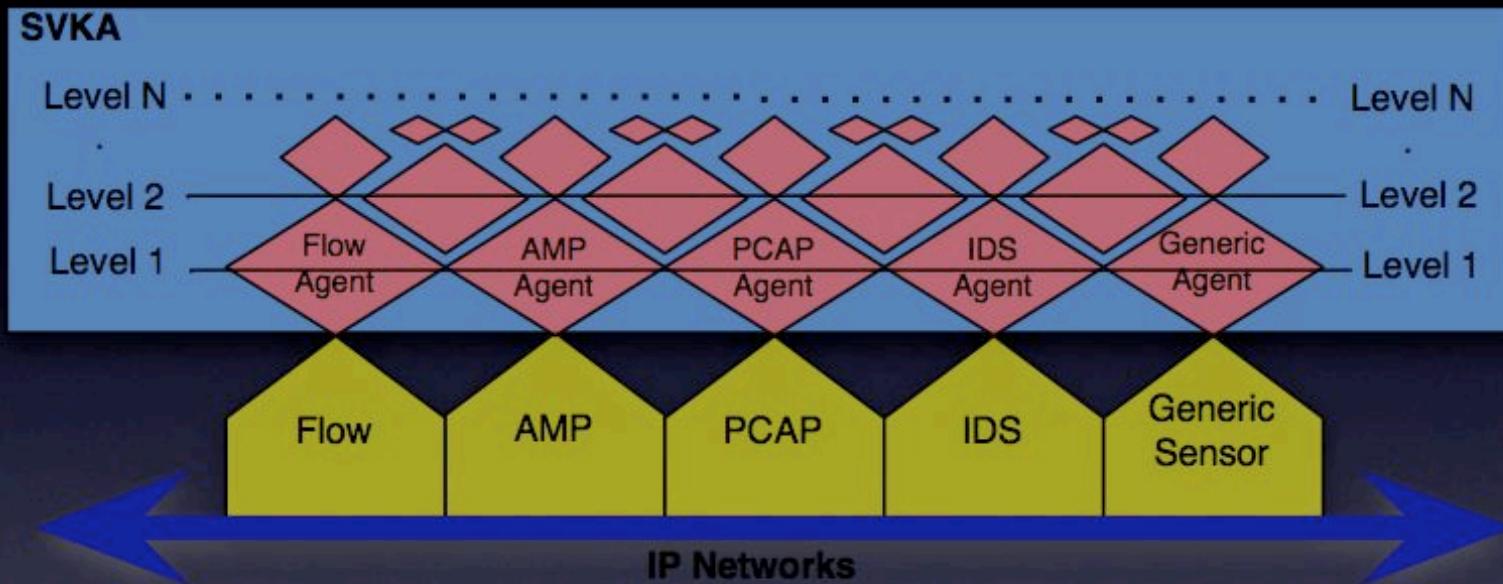
Flow Viewer Sensors



Flow Viewer Sensors



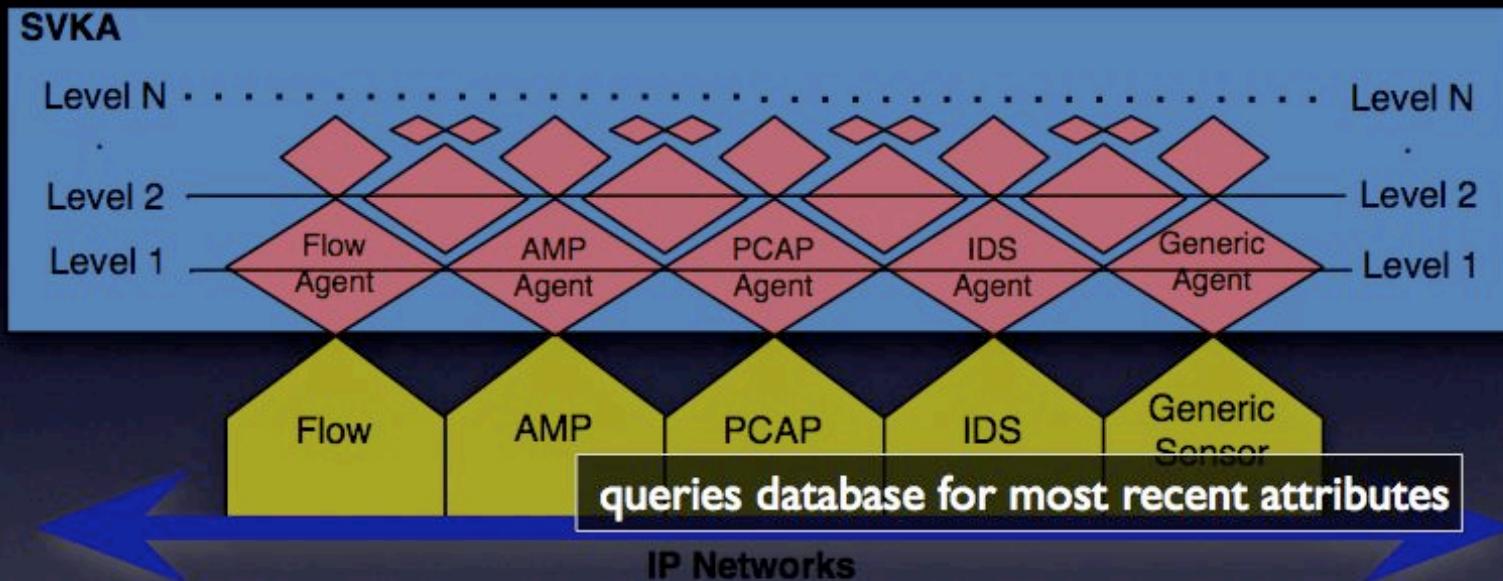
Flow Viewer Sensors



Flow Agent

consumes **rwf** & **rwcut** data

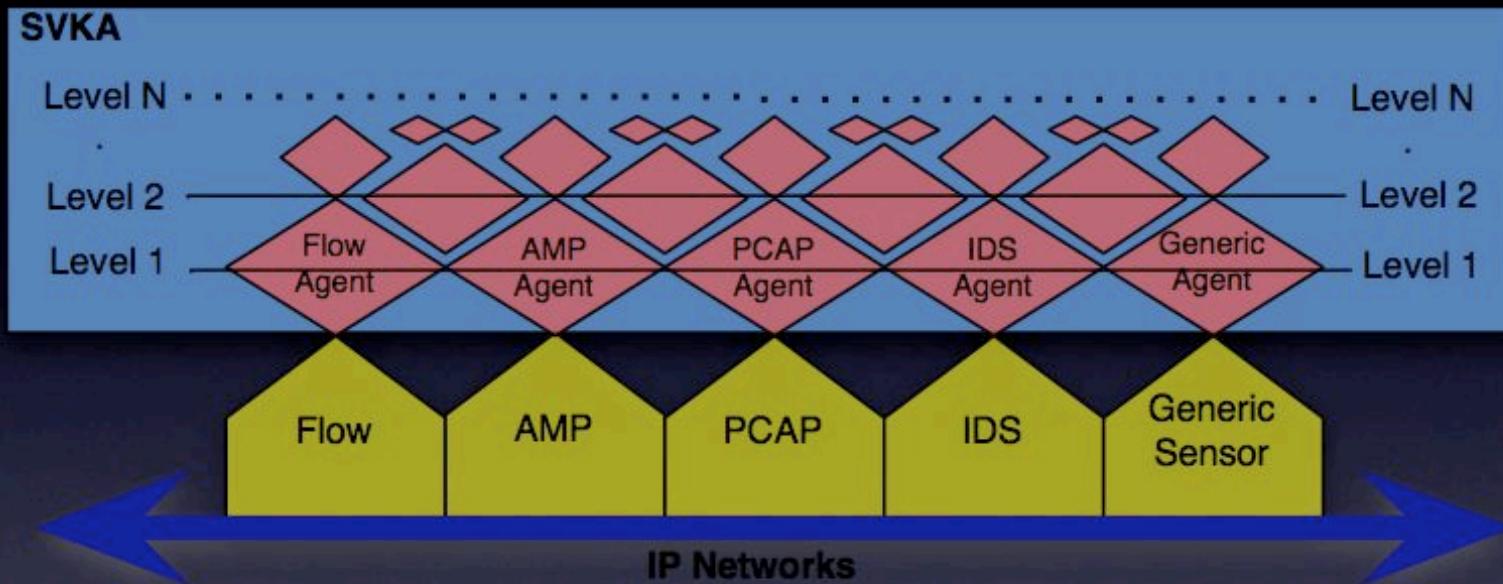
Flow Viewer Sensors



Flow Agent

consumes **rwf** & **rwcutf** data

Flow Viewer Sensors



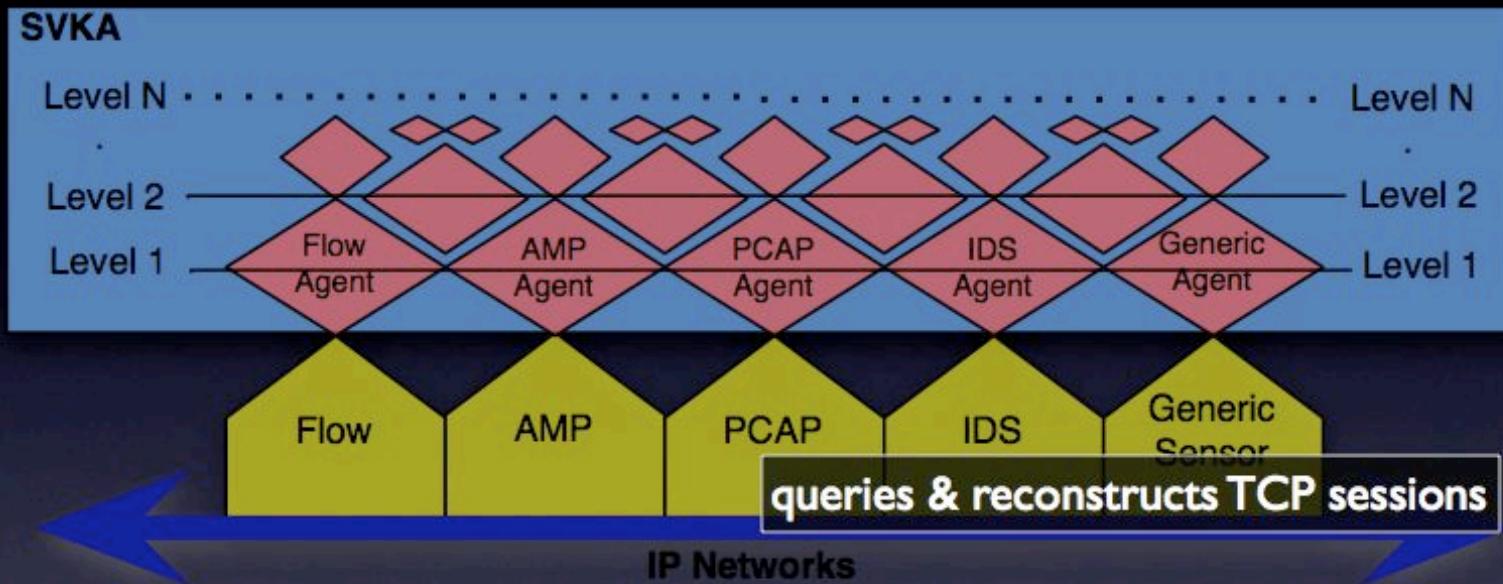
Flow Agent

consumes **rwf** & **rwcutf** data

AMP Agent

queries database for most recent attributes

Flow Viewer Sensors



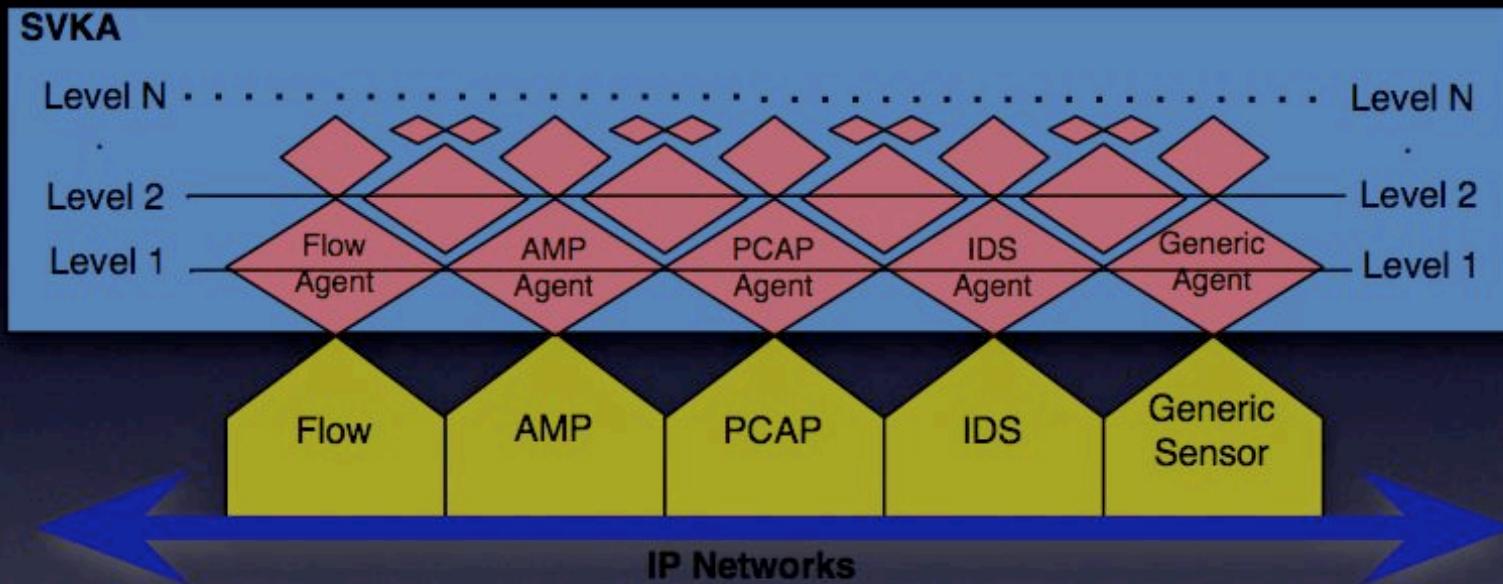
Flow Agent

consumes r wf & r wcut data

AMP Agent

queries database for most recent attributes

Flow Viewer Sensors



Flow Agent

consumes **rwf** & **rwcutf** data

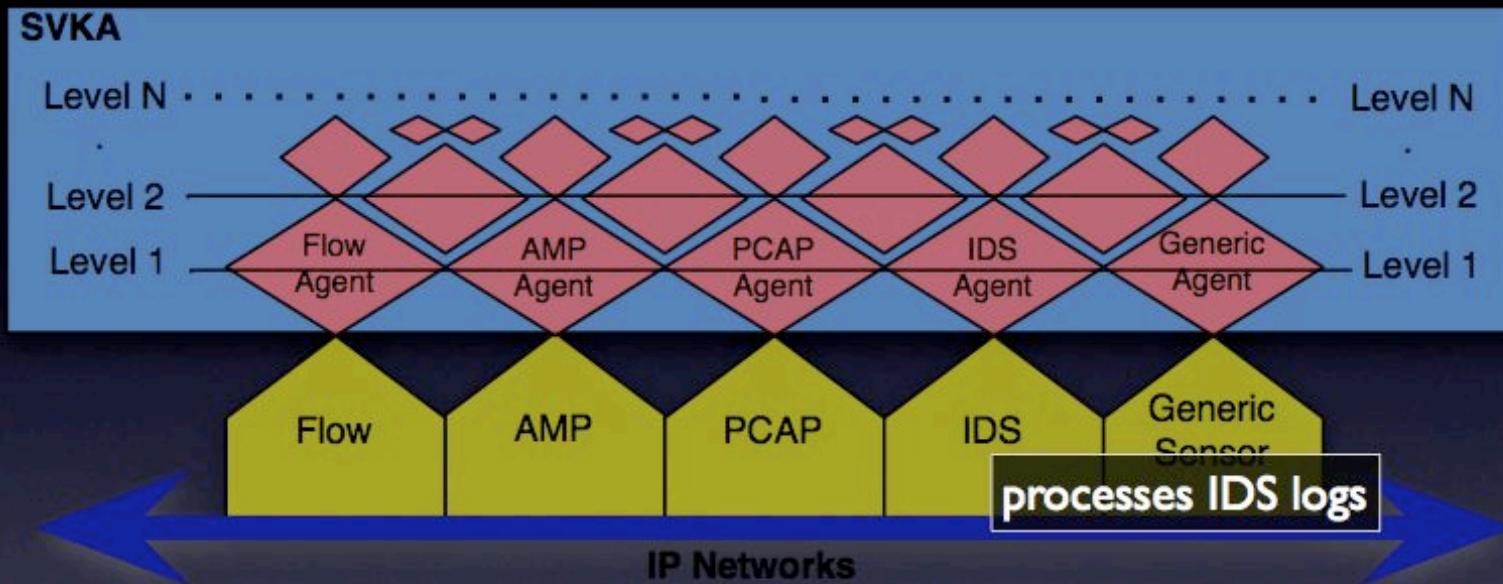
AMP Agent

queries database for most recent attributes

PCAP Agents

queries & reconstructs TCP sessions

Flow Viewer Sensors



Flow Agent

consumes **rwf** & **rwcutf** data

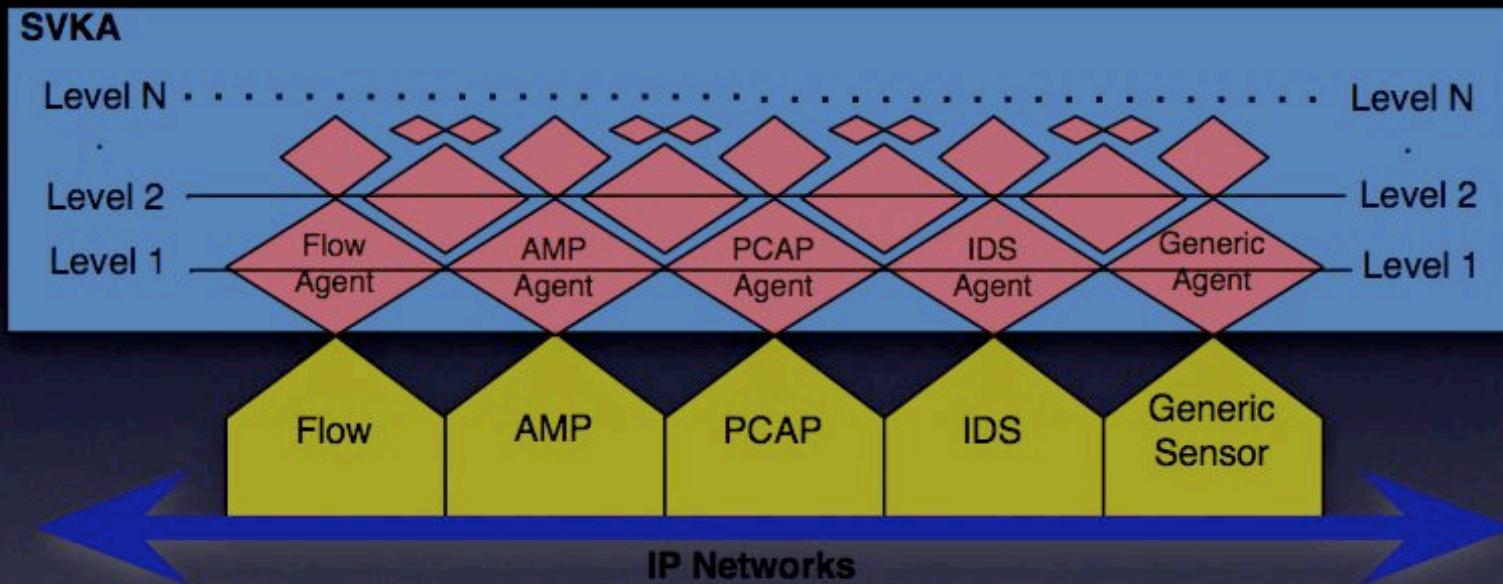
AMP Agent

queries database for most recent attributes

PCAP Agents

queries & reconstructs TCP sessions

Flow Viewer Sensors



Flow Agent

consumes **rwf** & **rwcut** data

AMP Agent

queries database for most recent attributes

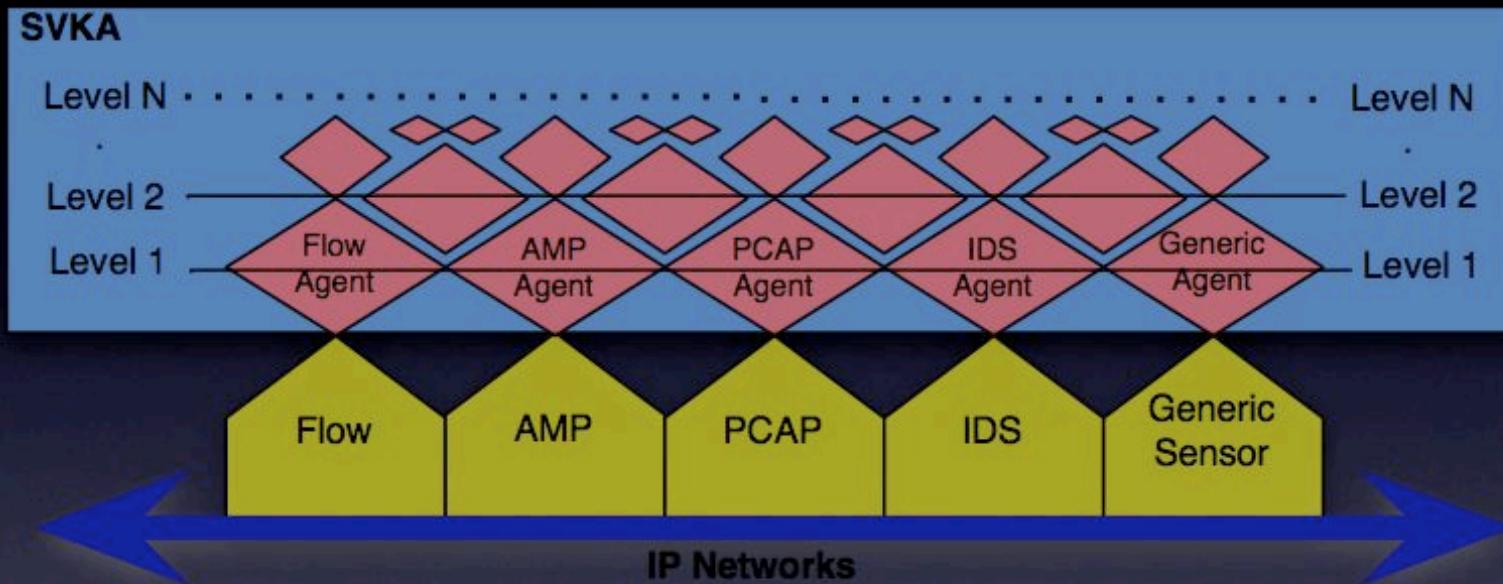
PCAP Agents

queries & reconstructs TCP sessions

IDS Agents

processes IDS logs

Flow Viewer Sensors



Flow Agent

consumes **rwf** & **rwcut** data



AMP Agent

queries database for most recent attributes



PCAP Agents

queries & reconstructs TCP sessions

IDS Agents

processes IDS logs

Flow Viewer Intelligent Agents

Flow Sensor

- Converts flow into ontology
- produces facts

AMP Agent

- uses correlations from Flow Agent
- query made on every unique IP seen
- produces visual events



Flow Agent

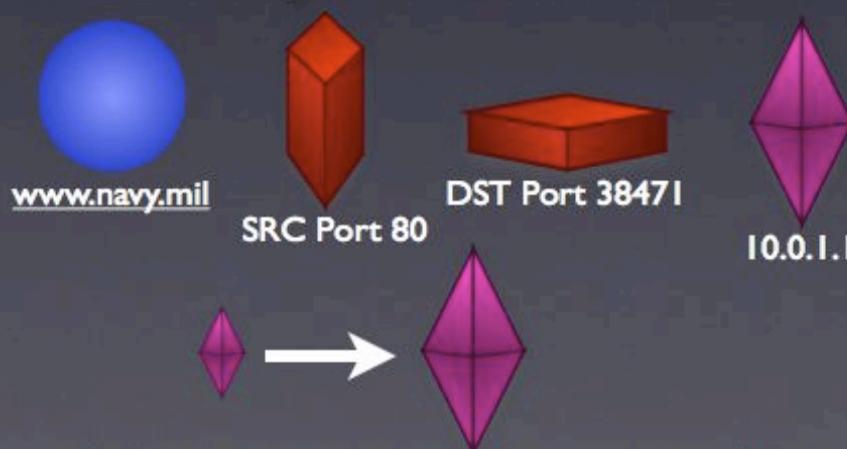
- correlates records
- counts and corroborates
- produces inferences
- produces visual events

Flow Viewer Visual Language

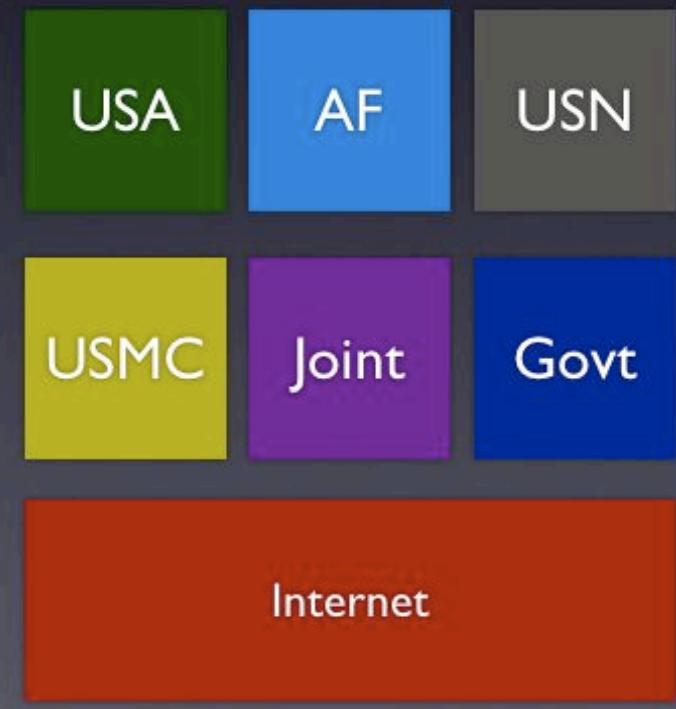
Leverage cultural knowledge



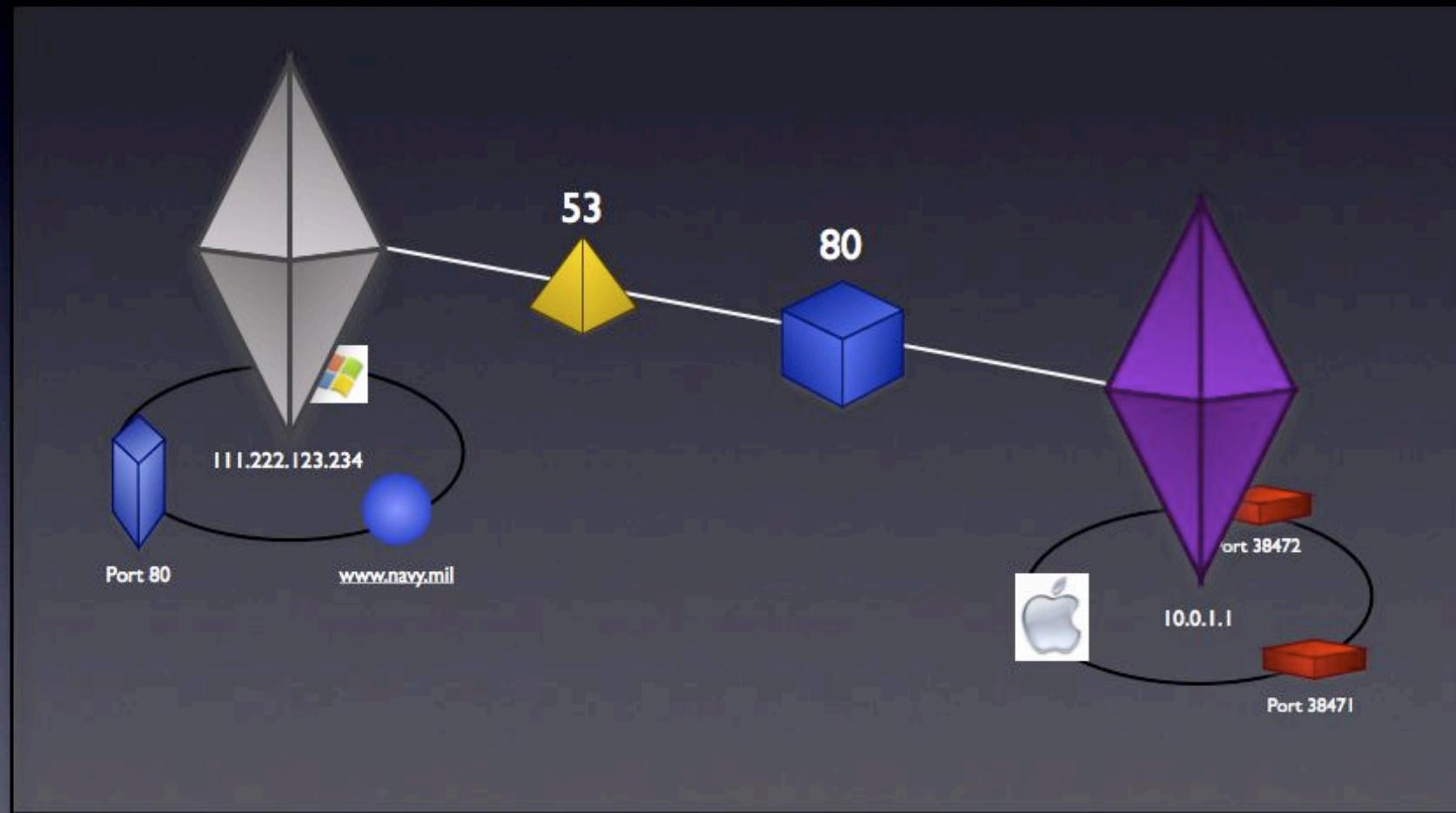
Use metaphors for abstract



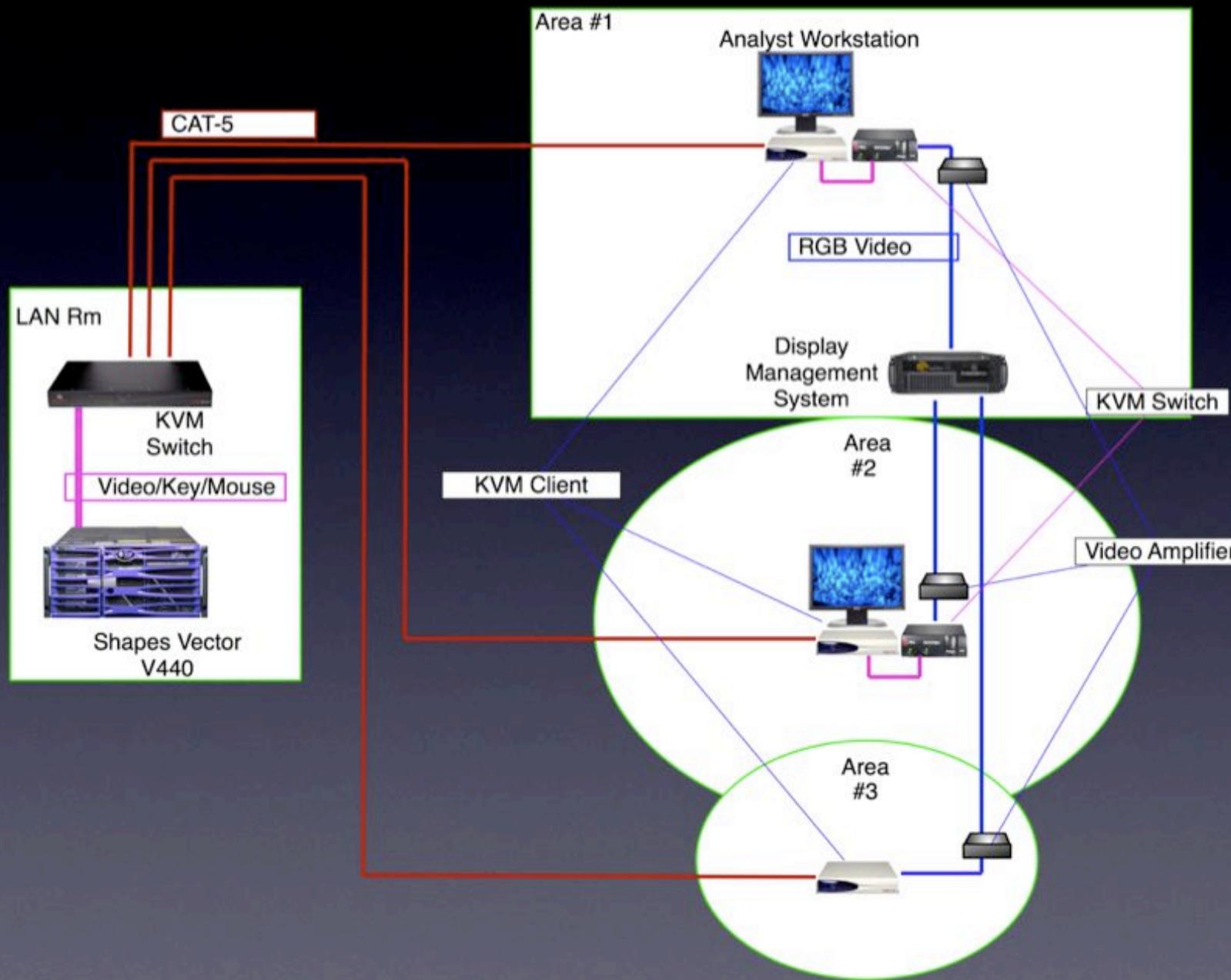
Color by ownership



Flow Viewer Visual Language



Test installation



Flow Viewer

Visualization

- **Tested** using:
 - 100-5000 nodes
 - 1M-3M flows
 - 10K-300K flows per hour
- Integrated **filtering** (rwfilter, SVKA filtering, visual filter)
- Visual ID
- **Queries**
- **Grouping** (e.g. domain, netblock, vulnerability)
- **Replay**-mode or Real-time
- Historic **visual context**
 - Replay 'on top of' known incident

Flow Viewer data prep

Include

- Incoming & outgoing
- Hub & core-to-core traffic
- Wide port ranges
- Time-span wider than the activity (minutes to hours)
- Suspect IPs and ranges

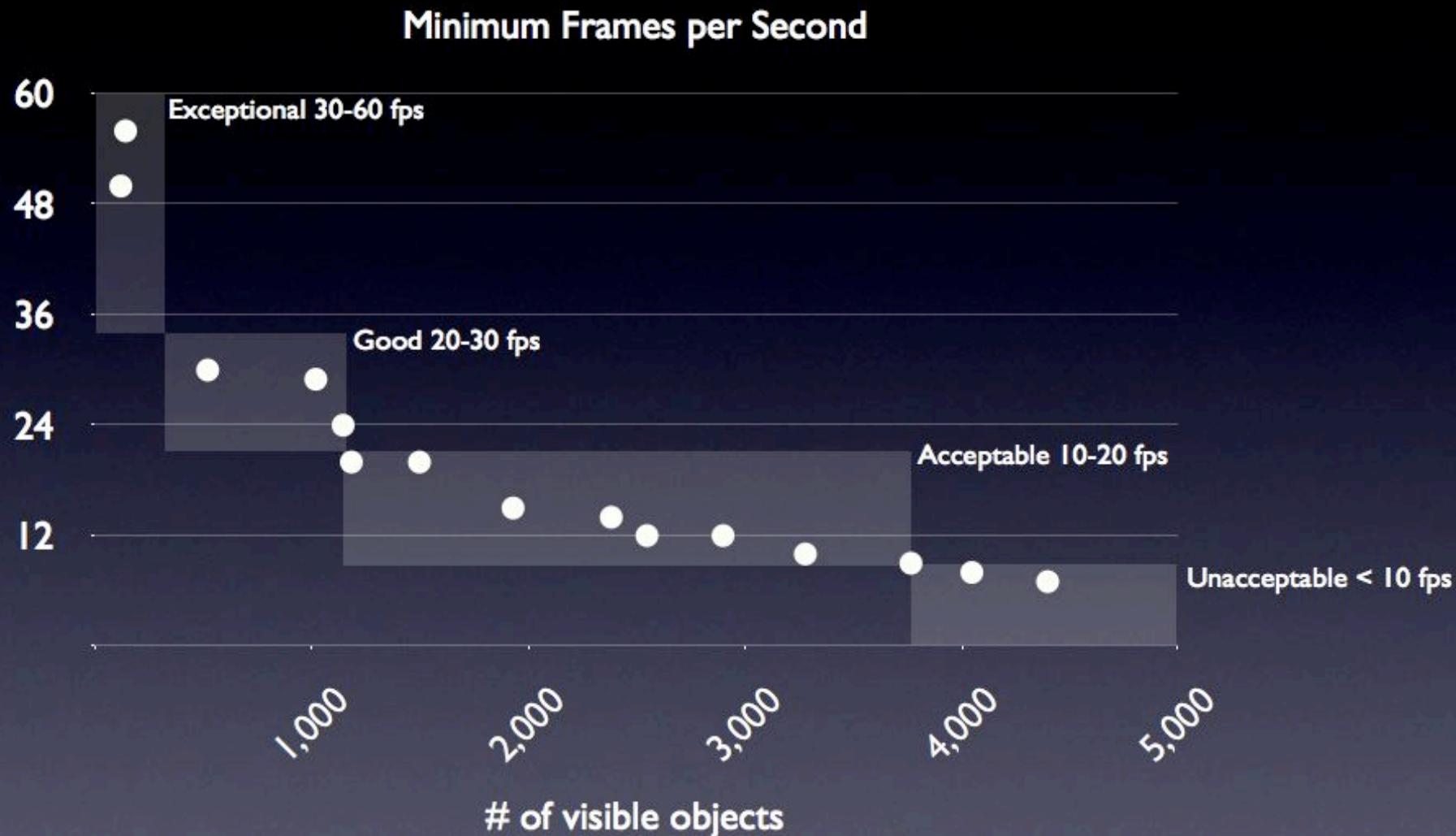
Filter

- Superfluous port traffic (e.g. 80, 53, 25)
- IPs that are unrelated to the incident

Sampling & Time

- Dense data
- Smear data across time resolution (~1 second)

Flow Viewer Performance



**Graphics performance on dual 1.5GHz SPARC SunFire v440 with Sun XVR 1200

Flow Viewer Performance

| Real-time Performance | Real-time Records / Hour | Optimal playback rate |
|-----------------------|--------------------------|-----------------------|
| Optimal | 10K-30K/hour | 10X Real-time |
| Acceptable | 40K-100K/hour | Real-time |
| Poor | 100K-300K/hour | 1/10 X Real-Time |

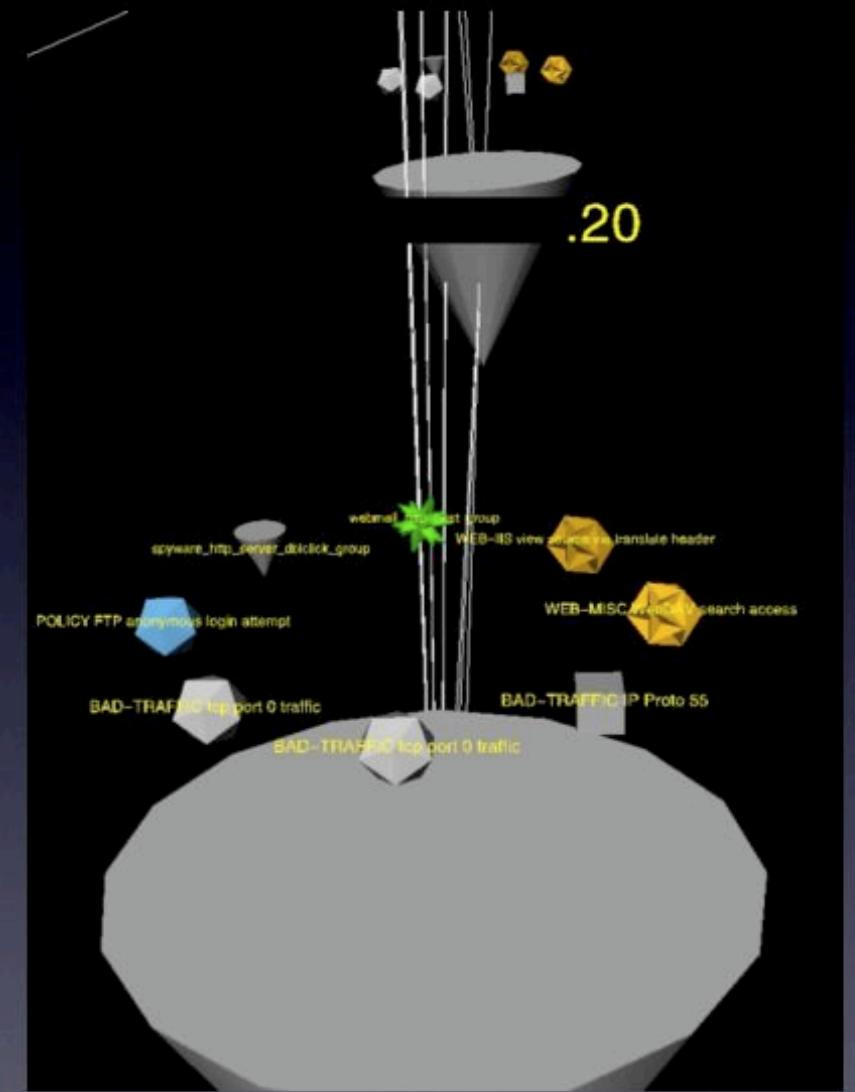
Sparse data sets can be viewed quickly
e.g. months of data in minutes

Dense data sets can be viewed slowly or filtered
e.g. seconds of data in minutes

Knowledge Depth vs Breadth

What trade-offs are we making?

- **UI Feedback?**
 - Haptic vs visual feedback
- **Data access?**
- **Random sequential access**
- **Training?**
 - Under-learned vs over-learned
 - Tool complexity
- **Meaning?**
 - Visual semantic vs text
 - Intuitive/iconic vs cryptic/coded



References

- [1] T. Abraham, Electronics, and Surveillance Research Laboratory (Australia). Information Technology Division. IDDM: Intrusion Detection Using Data Mining Techniques. DSTO Electronics and Surveillance Research Laboratory, 2001.
- [2] Mark Anderson, Dean Engelhardt, Damian Marriott, and Suneel Randhawa. Data processing and observation system, August 1 2006.
- [3] Mark Anderson, Dean Engelhardt, Damian Marriott, and Suneel Randhawa. Data view of a modelling system, April 11 2006.
- [4] H. H. Clark and W. G. Chase. On the process of comparing sentences against pictures. *Cognitive Psychology*, 3:472–517, 1972.
- [5] Herbert A. Colle and Gary B. Reid. The room effect: Metric spatial knowledge of local and separated regions. *Presence: Teleoperators and Virtual Environments*, 7(2):116–128, 1998.
- [6] Science Applications International Corporation. Intrusion Detection System System Protection Profile. National Security Agency, 9800 Savage Road, Fort Meade MD, 20755, version 1.4 edition, February 2002.
- [7] Stephen W. Draper and Donald A. Norman. *User Centered System Design: New Perspectives on Human-computer Interaction*. CRC, 1 edition, 1986.
- [8] D. Engelhardt and M. Anderson. A distributed multi-agent architecture for computer security situational awareness. *Information Fusion*, 2003. Proceedings of the Sixth International Conference of, 1, 2003.
- [9] Sunny Fugate. Visual language for tactical communication. In *Proceedings of the First Annual Visual and Iconic Language Conference*. SPAWAR Systems Center, San Diego, August 2007.
- [10] Sunny Fugate, Emily W. Medina, LorRaine Duffy, Dennis Magsombol, Omar Amezcuia, Gary Rogers, and Marion Ceruti. Next-generation tactical-situation-assessment technology (tsat): Iconic language. In Sunny Fugate, editor, *Proceedings of the First Annual Visual and Iconic Language Conference*. SPAWAR Systems Center, August 2007.
- [11] David Gamon and Allen D. Bragdon. *Brains That Work A Little Bit Differently: Recent Discoveries About Common Brain Diversities*. Barnes and Noble, 2000.
- [12] James K. Hahn, Hesham Fouad, Larry Gritz, and Jong Won Lee. Integrating sounds and motions in virtual environments. *Presence: Teleoperators and Virtual Environments*, 7(1):67–77, 1998.
- [13] T. Munzner. *INTERACTIVE VISUALIZATION OF LARGE GRAPHS AND NETWORKS*. PhD thesis, STANFORD UNIVERSITY, 2000.
- [14] Jakob Nielsen. *Usability Engineering (Interactive Technologies)*. Morgan Kaufmann, 1st edition, 1993.
- [15] CM Reed and NI Durlach. Short paper: Note on information transfer rates in human communication. *Presence: Teleoperators and Virtual Environments*, 7(5):509–518, 1998.
- [16] Walter Shepherd. *Shepherd's glossary of graphic signs and symbols*; Dent, London,, 1971.
- [17] Edward R. Tufte. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Graphics Press, Cheshire, Conn., 1997.
- [18] TS TULLIS. An evaluation of alphanumeric, graphic, and color information displays. *Human Factors*, 23:541–550, 1981.
- [19] D.J. Ward, A.F. Blackwell, and D.J.C. MacKay. Dasher—a data entry interface using continuous gestures and language models. *Proceedings of the 13th annual ACM symposium on User interface software and technology*, pages 129–137, 2000.
- [20] G.J. Wills. Nicheworks-interactive visualization of very large graphs. *Graph Drawing: 5th International Symposium, GD'97, Rome, Italy, September 18-20, 1997. Proceedings*, 1997.

Images

- [Jeff Han's Multi-Touch Screen Interface](http://creativecommons.org/licenses/by-sa/2.0/deed.en), Jeff Kubina, Flickr.com, license: <http://creativecommons.org/licenses/by-sa/2.0/deed.en>
- [Atari joystick](http://creativecommons.org/licenses/by-nc/2.0/deed.en), duncan, Flickr.com, license: <http://creativecommons.org/licenses/by-nc/2.0/deed.en>
- [Headphones](http://creativecommons.org/licenses/by-sa/2.0/deed.en), daxtoor, Flickr.com, license: <http://creativecommons.org/licenses/by-sa/2.0/deed.en>



Next Generation Tactical Situation Assessment Technology (NG-TSAT)



Objective: Next-generation Tactical Chat. Icon-based situation assessment (SA) language supported by wireless gesture-recognition gloves used in hostile or noisy (silence-mandated) environments

Description of Effort:

- 1. Linguistic Analysis:** Analysis of current C² chat logs to determine speech patterns and repetitive SA concepts/themes
- 2. Iconic Language Development:** Output of linguistic analysis determines candidate icons representing most prevalent SA "themes;" development of prototype C² iconic SA language
- 3. Wireless, Gesture-Recognition Gloves:** Develop wireless gloves that recognize C² icons/gestures which can transmit across network to distributed warfighters (replacing keyboard input when in MOPP)

Benefits of TSAT:

Compressed Chat (25% ↓ content; 50% ↓ reduction in production time) for rapid SA dissemination.
Gesture-recognition in very noisy, distributed ops, or in very austere environments (e.g., the moon)
Challenges:

1. No current method or theory for chat-meaning compression; currently done in prose; computer linguistic analysis of unstructured text still neoteric.
2. Wireless gesture recognition glove technology still in infant stages of development; focused on commercial animation support, not on disciplined language support

TRL: Chat: TRL 1-2; Gesture-recognition: TRL 1-4



Major Milestones FY06:

Linguistic analysis discovery of common C² SA themes
Development of icon/symbols for candidate SA themes
Development of proof-of-concept wireless gesture-recognition glove

Period of Performance: 2007-2012

PI contact info: Dr. LorRaine Duffy, (619) 553-9222,
LorRaine.Duffy@navy.mil, SSC San Diego, CA

Synesthesia

Synesthesia: "a neurological condition in which two or more senses are coupled."

"loud color" "sharp laugh" "bitter wind"

grapheme color synesthesia - letters or numbers are perceived as inherently **colored**

How many numbers contain the digit 6?

| | | | | | | |
|------|------|------|------|------|------|------|
| 9910 | 9972 | 3292 | 7602 | 82 | 9054 | |
| 5636 | 2710 | 1944 | 6330 | 6560 | 8101 | |
| 5177 | 1955 | 7029 | 4083 | 4643 | 5710 | |
| 4935 | 2256 | 1495 | 1025 | 8375 | 8518 | |
| 80 | 797 | 2610 | 3008 | 8784 | 1854 | 2383 |
| 9728 | 4523 | 573 | 5914 | 7975 | 281 | |
| 6664 | 2682 | 7689 | 7753 | 273 | 5597 | |
| 799 | 9960 | 1437 | 4534 | 8601 | 4563 | |
| 6734 | 647 | 9409 | 6543 | 4827 | 2398 | |
| 1532 | | | | | | |

Is this easier?

| | | | | | | |
|------|------|------|------|------|------|------|
| 9910 | 9972 | 3292 | 7602 | 82 | 9054 | 5636 |
| 2710 | 1944 | 6330 | 6560 | 8101 | 5177 | |
| 1955 | 7029 | 4083 | 4643 | 5710 | 4935 | |
| 2256 | 1495 | 1025 | 8375 | 8518 | 80 | 797 |
| 2610 | 3008 | 8784 | 1854 | 2383 | 9728 | |
| 4523 | 573 | 5914 | 7975 | 281 | 6664 | 2682 |
| 7689 | 7753 | 273 | 5597 | 799 | 9960 | 1437 |
| 4534 | 8601 | 4563 | 6734 | 647 | 9409 | |
| 6543 | 4827 | 2398 | 1532 | | | |

Emulating Synesthesia

These methods can be used achieve
sequence disambiguation and

| | | | | | | |
|------|------|------|------|------|------|------|
| 9910 | 9972 | 3292 | 7602 | 82 | 9054 | |
| 5636 | 2710 | 1944 | 6330 | 6560 | 8101 | |
| 5177 | 1955 | 7029 | 4083 | 4643 | 5710 | |
| 4935 | 2256 | 1495 | 1025 | 8375 | 8518 | |
| 80 | 797 | 2610 | 3008 | 8784 | 1854 | 2383 |
| 9728 | 4523 | 573 | 5914 | 7975 | 281 | |
| 6664 | 2682 | 7689 | 7753 | 273 | 5597 | |
| 799 | 9960 | 1437 | 4534 | 8601 | 4563 | |
| 6734 | 647 | 9409 | 6543 | 4827 | 2398 | |
| 1532 | | | | | | |

| | | | | | | |
|------|------|------|------|------|------|------|
| 9910 | 9972 | 3292 | 7602 | 82 | 9054 | |
| 5636 | 2710 | 1944 | 6330 | 6560 | 8101 | |
| 5177 | 1955 | 7029 | 4083 | 4643 | 5710 | |
| 4935 | 2256 | 1495 | 1025 | 8375 | 8518 | |
| 80 | 797 | 2610 | 3008 | 8784 | 1854 | 2383 |
| 9728 | 4523 | 573 | 5914 | 7975 | 281 | |
| 6664 | 2682 | 7689 | 7753 | 273 | 5597 | |
| 799 | 9960 | 1437 | 4534 | 8601 | 4563 | |
| 6734 | 647 | 9409 | 6543 | 4827 | 2398 | |
| 1532 | | | | | | |

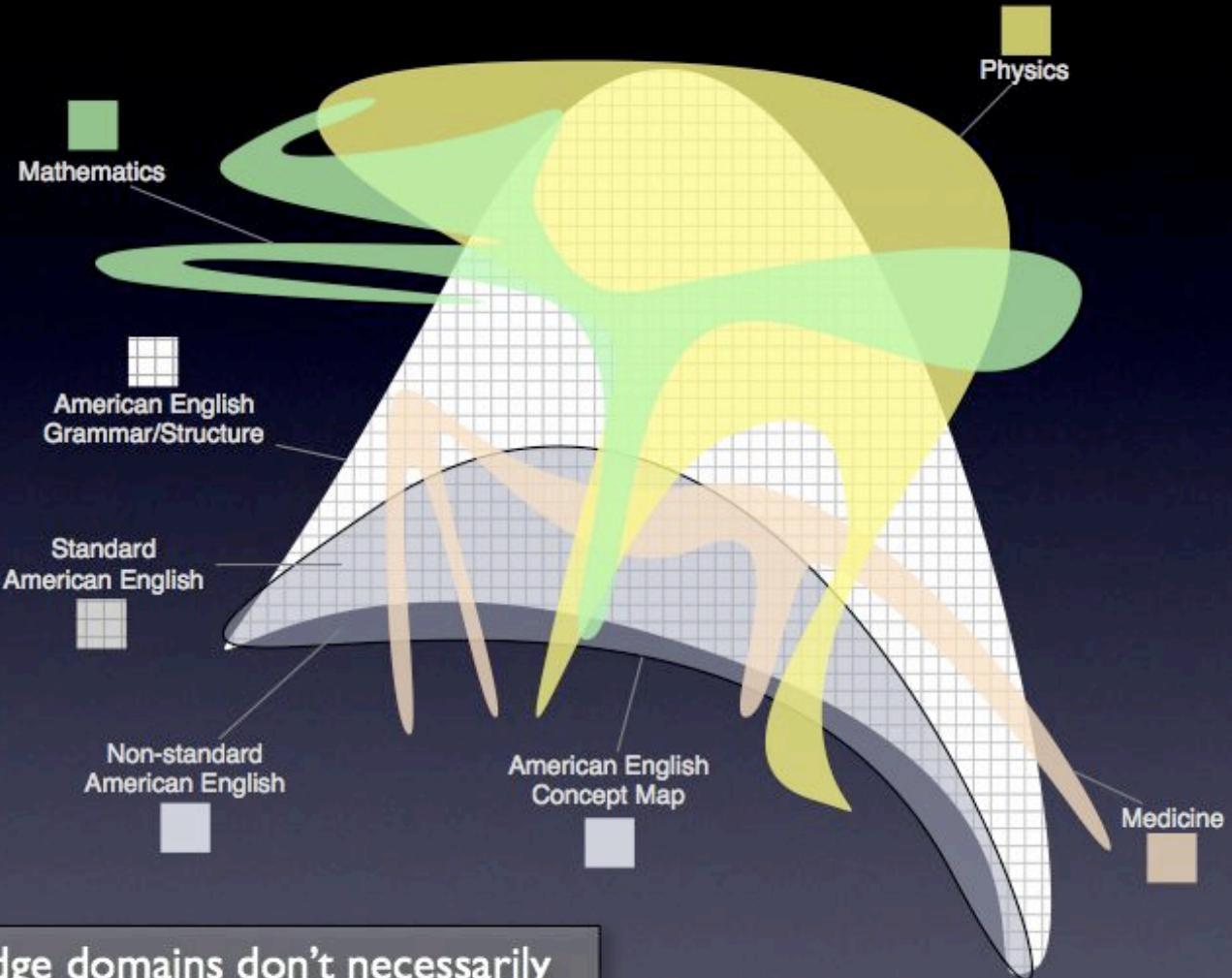
Emulating Synesthesia

192.168.1.232
129.168.1.233

192.168.1.232
129.168.1.233

192.168.1.232
129.168.1.233

Language Domains



Cultures and knowledge domains don't necessarily use the same lexicon or even the same grammar!

How does the CND lexicon map to common language?
Technical language? Military/tactical language?