Monitoring Data Fusion for Intrusion Tolerance

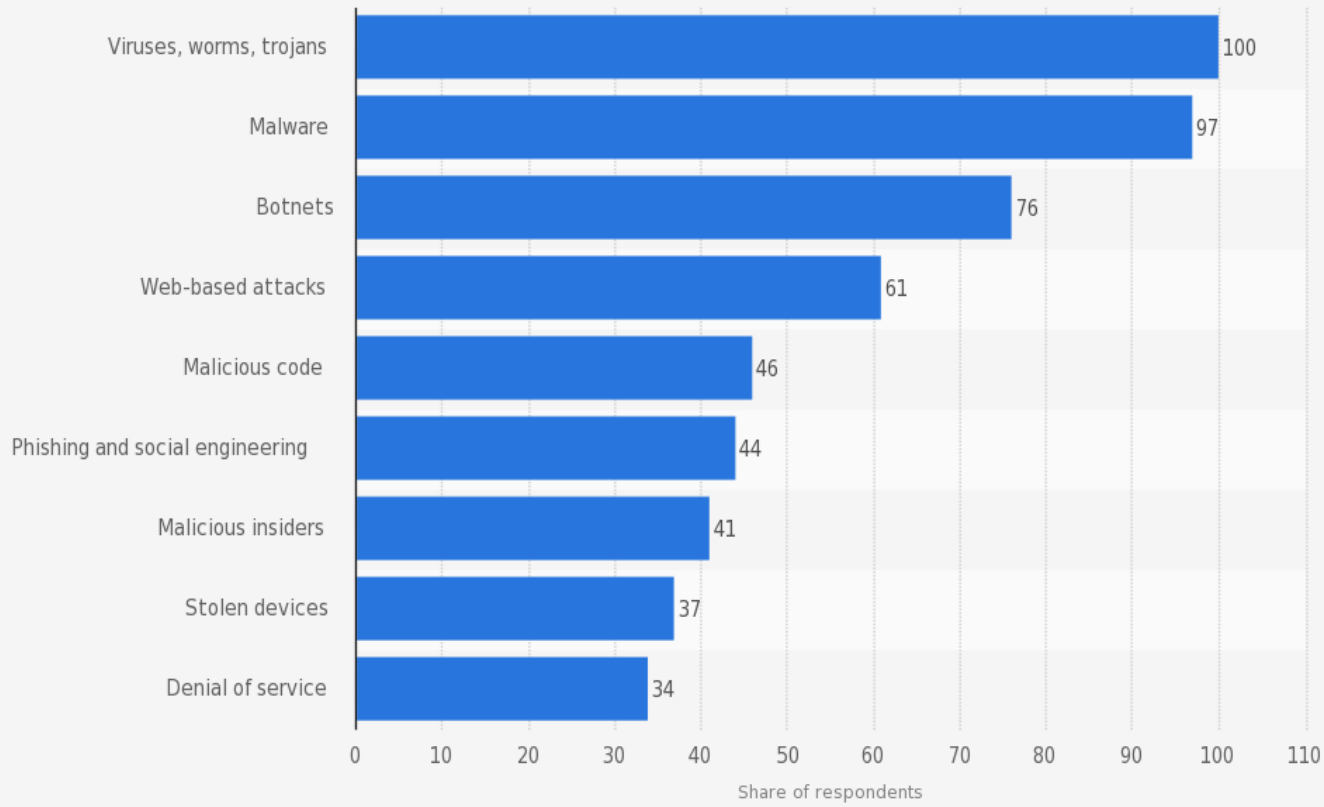
Atul Bohara
P.I.: W. H. Sanders
ACC Seminar, Oct 7, 2015

Overview

- Protect a real-world networked system against malicious activities
  - E.g., Enterprise network / campus network / cloud data center
  - Prevention techniques are not sufficient
  - Need to rely on security monitoring and detection

- **Questions**
  - How to make *overall sense* out of the data generated by these monitors?
  - How to support intrusion detection and automatic response?

Motivation



Quellen::
Ponemon Institute; Hewlett-Packard (HP Enterprise Security)
© Statista 2015

Additional Information:
United States; Ponemon Institute; June 2014; 59 U.S. organizations

statista

Motivation

- **Security Breaches are increasing**
- **Security data is large and unmanageable**

Problem

- System sizes are growing
  - New attack types
  - Variety of tools and techniques required to protect the system
- Overwhelmingly big and heterogeneous security data
- Difficult to have a holistic view of system

Can we do something?

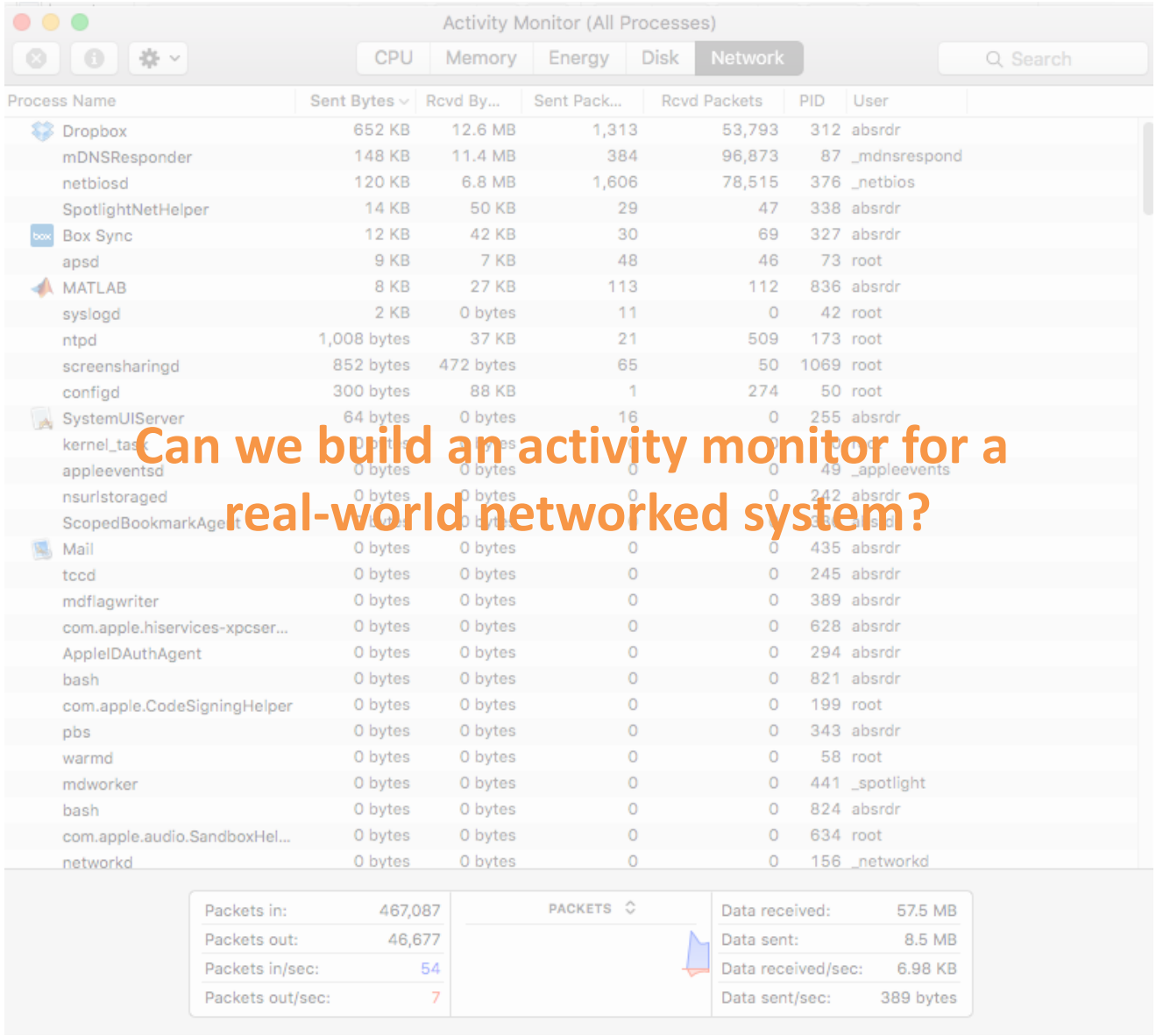
Activity Monitor (All Processes)

CPU | Memory | Energy | Disk | Network | Search

Process Name	Sent Bytes	Rcvd By...	Sent Pack...	Rcvd Packets	PID	User
Dropbox	652 KB	12.6 MB	1,313	53,793	312	absrdr
mDNSResponder	148 KB	11.4 MB	384	96,873	87	_mdnsrespond
netbiosd	120 KB	6.8 MB	1,606	78,515	376	_netbios
SpotlightNetHelper	14 KB	50 KB	29	47	338	absrdr
Box Sync	12 KB	42 KB	30	69	327	absrdr
apsd	9 KB	7 KB	48	46	73	root
MATLAB	8 KB	27 KB	113	112	836	absrdr
syslogd	2 KB	0 bytes	11	0	42	root
ntpd	1,008 bytes	37 KB	21	509	173	root
screensharingd	852 bytes	472 bytes	65	50	1069	root
configd	300 bytes	88 KB	1	274	50	root
SystemUIServer	64 bytes	0 bytes	16	0	255	absrdr
kernel_task	0 bytes	0 bytes	0	0	0	root
appleeventsd	0 bytes	0 bytes	0	0	49	_appleevents
nsurlstoraged	0 bytes	0 bytes	0	0	242	absrdr
ScopedBookmarkAgent	0 bytes	0 bytes	0	0	386	absrdr
Mail	0 bytes	0 bytes	0	0	435	absrdr
tccd	0 bytes	0 bytes	0	0	245	absrdr
mdflagwriter	0 bytes	0 bytes	0	0	389	absrdr
com.apple.hiservices-xpcser...	0 bytes	0 bytes	0	0	628	absrdr
AppleIDAuthAgent	0 bytes	0 bytes	0	0	294	absrdr
bash	0 bytes	0 bytes	0	0	821	absrdr
com.apple.CodeSigningHelper	0 bytes	0 bytes	0	0	199	root
pbs	0 bytes	0 bytes	0	0	343	absrdr
warmd	0 bytes	0 bytes	0	0	58	root
mdworker	0 bytes	0 bytes	0	0	441	_spotlight
bash	0 bytes	0 bytes	0	0	824	absrdr
com.apple.audio.SandboxHel...	0 bytes	0 bytes	0	0	634	root
networkd	0 bytes	0 bytes	0	0	156	_networkd

Packets in:	467,087	PACKETS	Data received:	57.5 MB
Packets out:	46,677		Data sent:	8.5 MB
Packets in/sec:	54		Data received/sec:	6.98 KB
Packets out/sec:	7		Data sent/sec:	389 bytes

Can we do something?



Can we build an activity monitor for a real-world networked system?

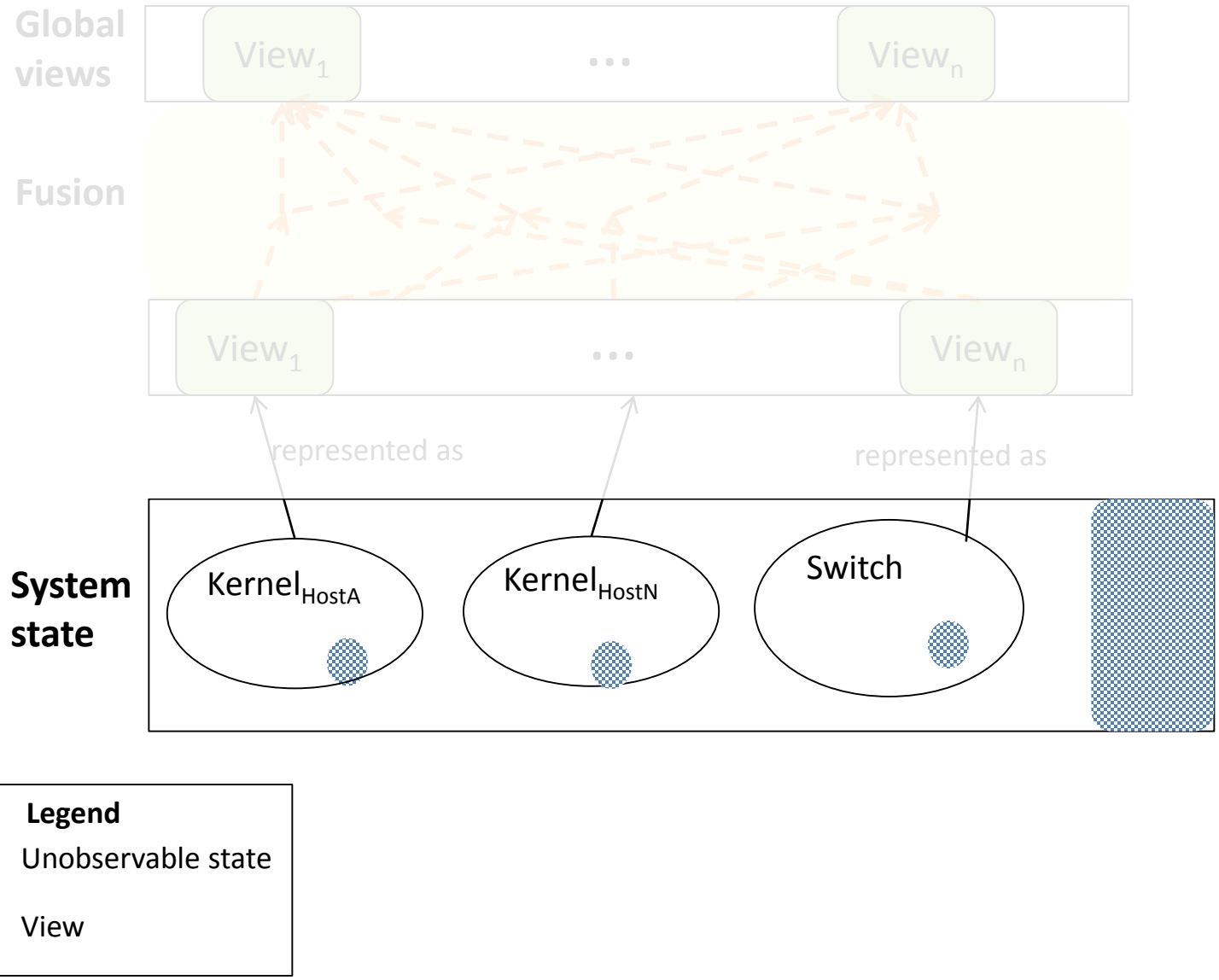
Our Approach

- Learn profile of the system over time
  - Define important categories of information
  - Host, network, users, application
  - Define multiple *views* of the system
- Generate and maintain these views in real-time
- Ability to drill-down and roll-up

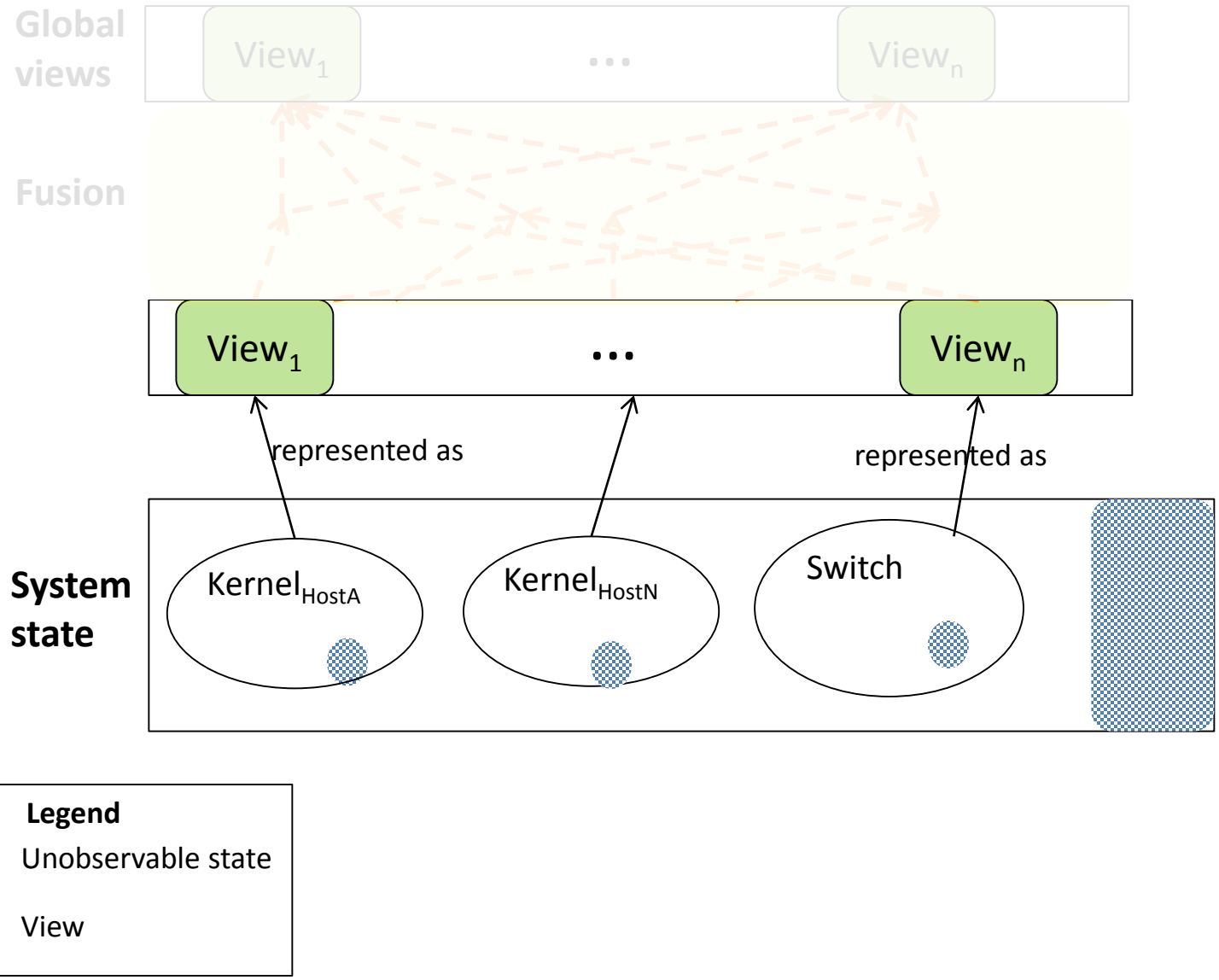
Monitoring Fusion: create, maintain and present higher-level system views.

Detect anomalies, policy violations; act upon them

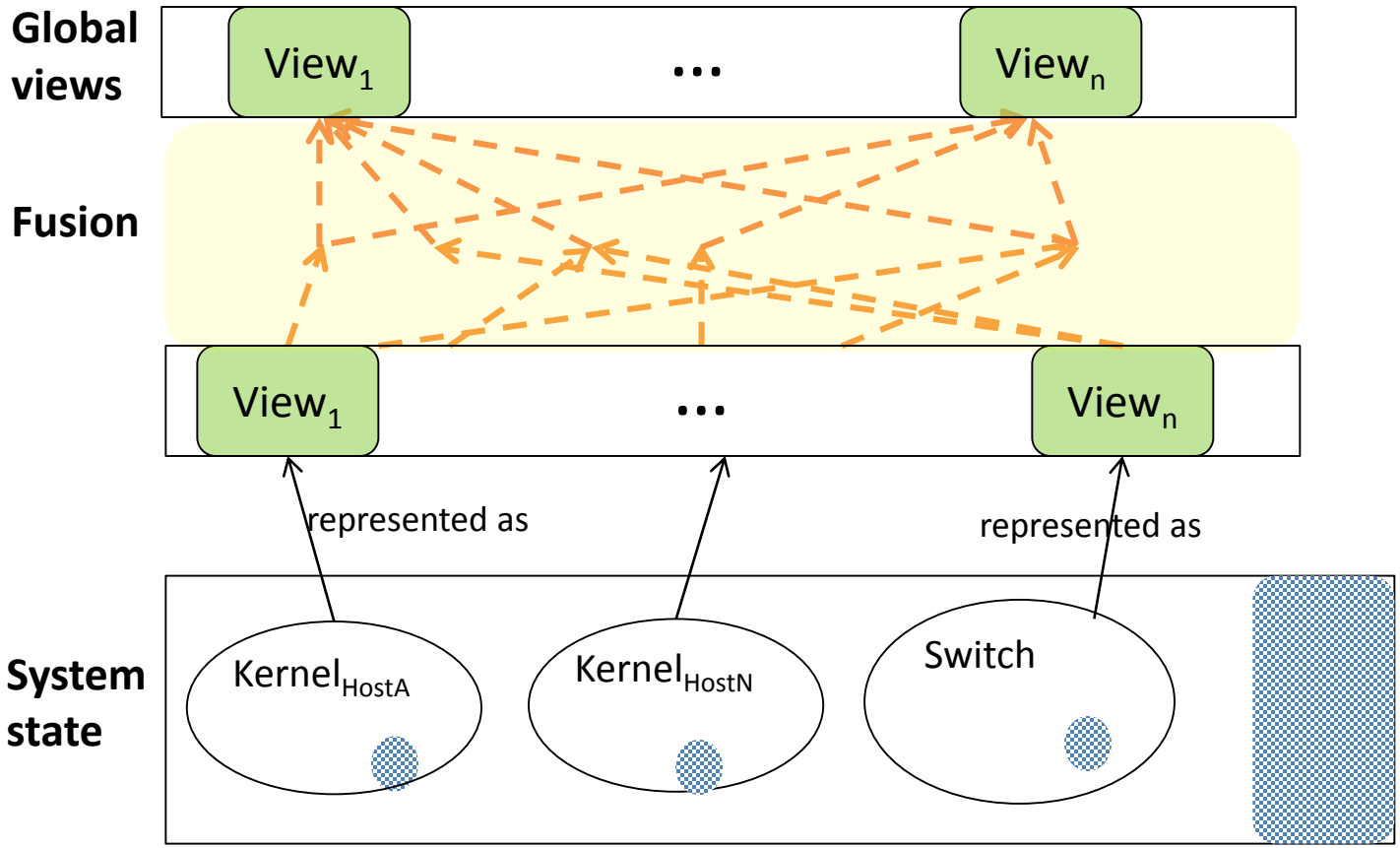
System Views

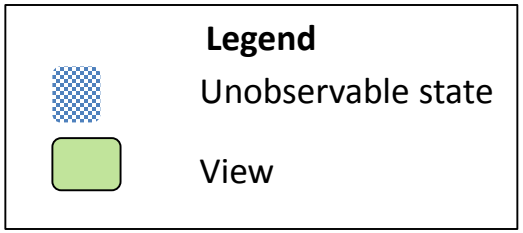


System Views



System Views





Outline

1. Generating System Views
    1. Handling Big and heterogeneous data

2. Issues
    1. Conflicting evidences
    2. Independence of data sources
    3. Ensure security of the procedure

Outline

1. **Generating System Views**
    1. **Handling Big and heterogeneous data**

2. Issues
    1. Conflicting evidences
    2. Independence of data sources
    3. Ensure security of the procedure

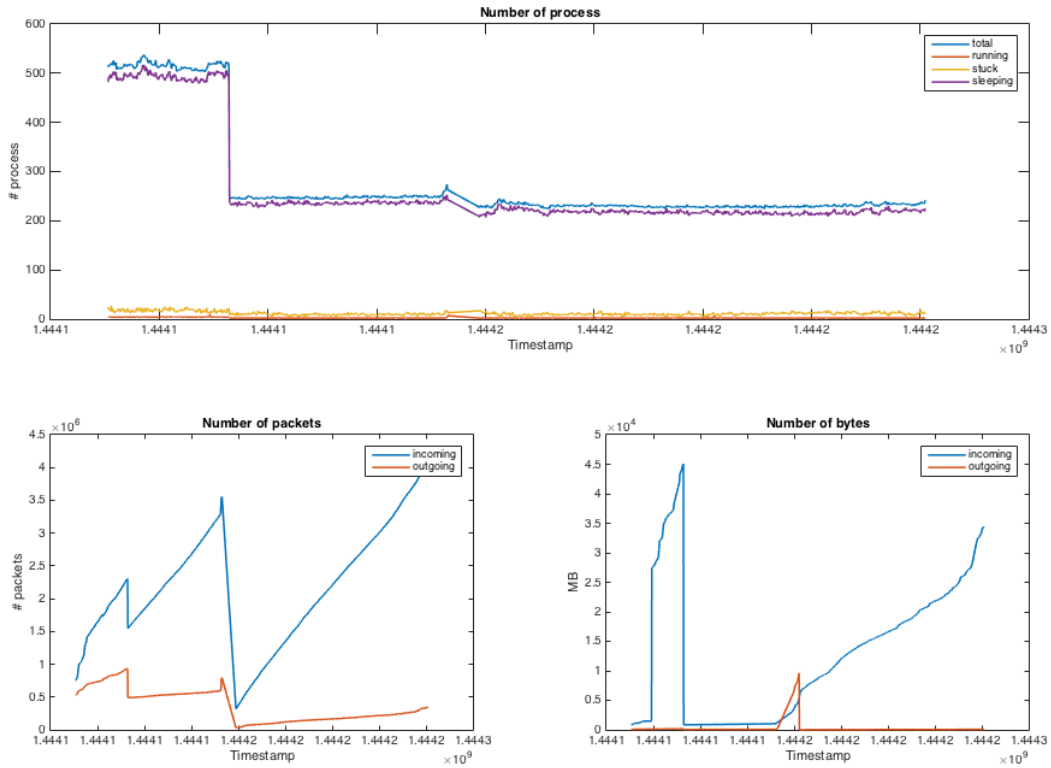
Generating System Views

- Define a catalog of views
- Combine data from multiple sources intelligently
  - Monitor selection and data extraction
  - Data reduction techniques
  - Hierarchical and on-demand fusion
- Group parts of system intelligently
  - Topological grouping
  - Grouping based on behavior

Example 1: Process-Traffic View

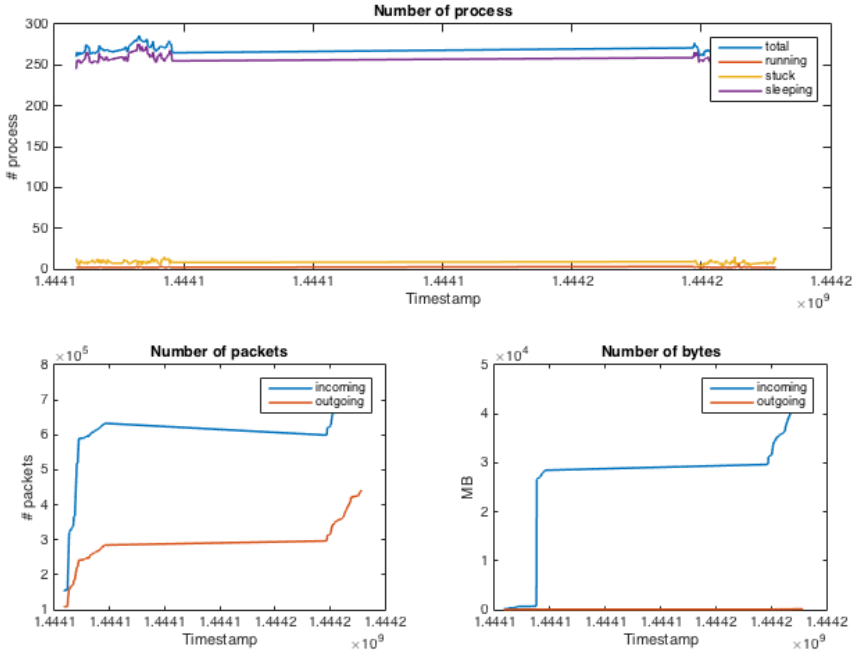
Number of processes, network bytes, and network packets

Combined profile of a group of two hosts

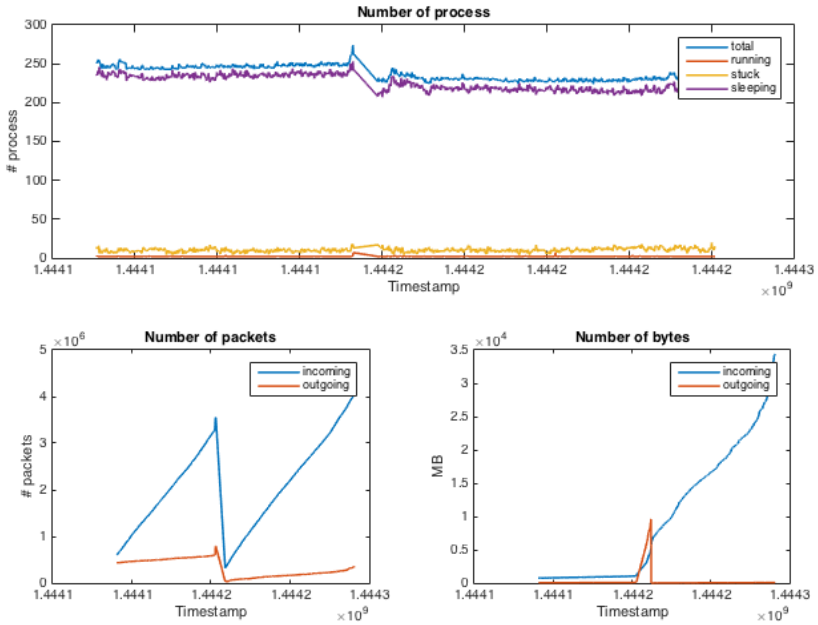


Drilling-down in Process-Traffic View

Host 1

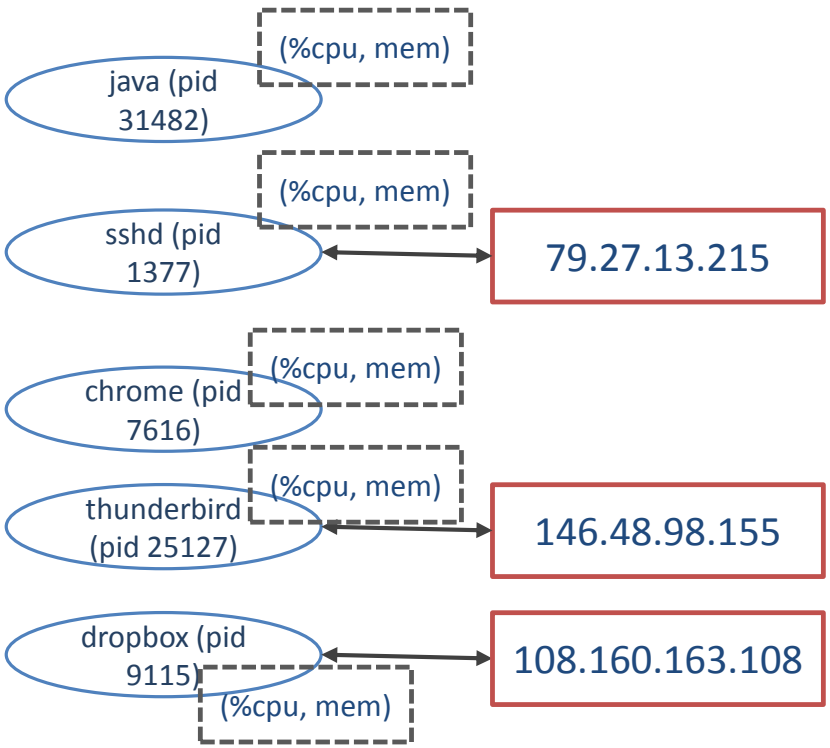


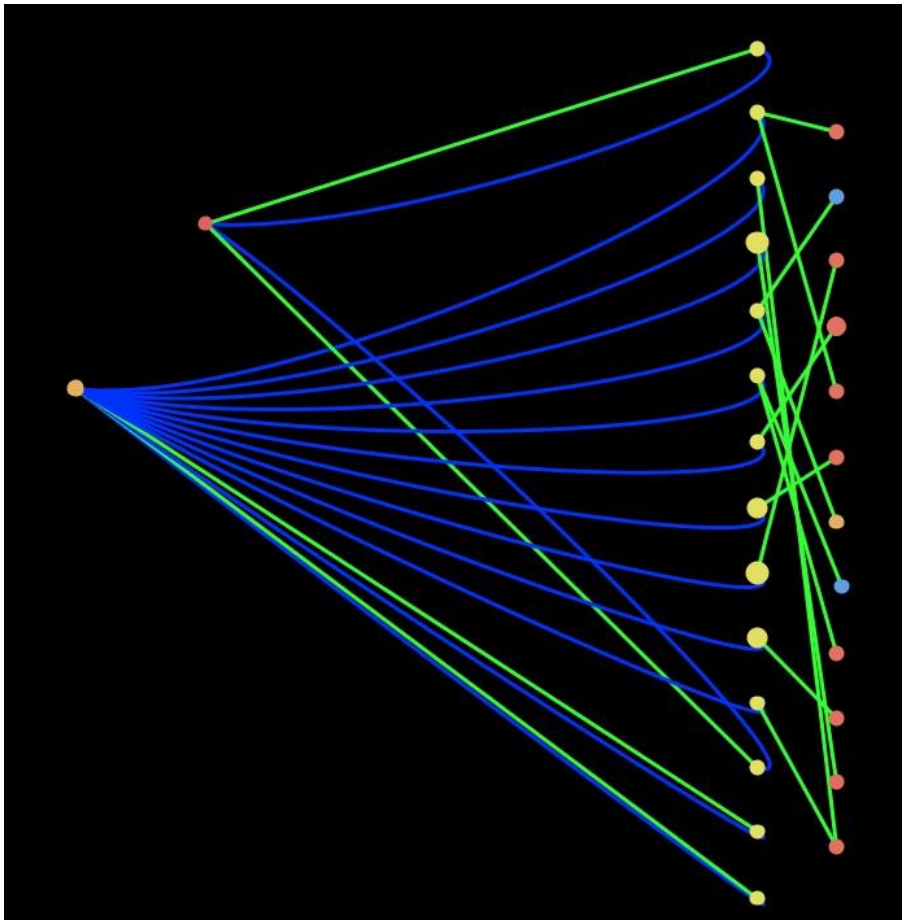
Host 2



Example 2: Network-Resources View

Local resource usage on hosts linked to every network activity





Raw Monitoring Data

mongod ... java screen

```
lsofMac===1430858964===1429669761===Box\x20Sync===44859===absrdr===60u===IPv4===0x501b4d6018193ab1===0t0==
=TCP===CLOSE_WAIT===130.126.138.124===54335===166.78.117.148===443
lsofMac===1430858964===1429669761===Box\x20Sync===44859===absrdr===61u===IPv4===0x501b4d6015da8431===0t0==
=TCP===ESTABLISHED===130.126.138.124===54499===74.112.184.85===443
lsofMac===1430858964===1429669761===Box\x20Sync===44859===absrdr===65u===IPv4===0x501b4d6017862ab1===0t0==
=TCP===ESTABLISHED===130.126.138.124===54498===74.112.184.85===443
lsofMac===1430858964===1429669761===Box\x20Sync===44859===absrdr===67u===IPv4===0x501b4d6015553041===0t0==
=TCP===ESTABLISHED===130.126.138.124===54492===74.112.184.85===443
lsofMac===1430858964===1429669761===Box\x20Sync===44859===absrdr===69u===IPv4===0x501b4d60178d1771===0t0==
=TCP===CLOSE_WAIT===130.126.138.124===50650===74.112.185.96===443
lsofMac===1430858964===1429669761===Box\x20Sync===44859===absrdr===71u===IPv4===0x501b4d60197bf771===0t0==
=TCP===CLOSE_WAIT===130.126.138.124===50649===74.112.184.85===443
lsofMac===1430858964===1429669761===Box\x20Sync\x20===44874===absrdr===26u===IPv4===0x501b4d600c3771e1===0
t0===TCP===CLOSE_WAIT===130.126.138.124===60207===74.112.184.85===443
lsofMac===1430858964===1429669761===Box\x20Sync\x20===44874===absrdr===27u===IPv4===0x501b4d60030f2911===0
t0===TCP===CLOSE_WAIT===130.126.138.124===60205===74.112.184.85===443
lsofMac===1430858964===1429669761===Box\x20Sync\x20===44874===absrdr===35u===IPv4===0x501b4d60197beea1===0
t0===TCP===CLOSE_WAIT===130.126.138.124===60206===74.112.184.85===443
lsofMac===1430858964===1429669761===java===45474===absrdr===46u===IPv6===0x501b4d6015d6cc61===0t0===TCP===
ESTABLISHED===127.0.0.1===2181===127.0.0.1===61436
lsofMac===1430858964===1429669761===java===45477===absrdr===44u===IPv6===0x501b4d6015d6d161===0t0===TCP===
ESTABLISHED===127.0.0.1===61436===127.0.0.1===2181
lsofMac===1430858964===1429669761===Safari===49970===absrdr===45u===IPv4===0x501b4d6015554ab1===0t0===TCP=
==ESTABLISHED===130.126.138.124===54487===173.194.77.106===443
lsofMac===1430858964===1429669761===Safari===49970===absrdr===46u===IPv4===0x501b4d6015554ab1===0t0===TCP=
==ESTABLISHED===130.126.138.124===54487===173.194.77.106===443
lsofMac===1430858964===1429669761===com.apple.WebKi===49972===absrdr===8u===IPv4===0x501b4d60181c8d01===0t
0===TCP===ESTABLISHED===130.126.138.124===54510===173.222.4.10===443
lsofMac===1430858964===1429669761===com.apple.WebKi===49972===absrdr===10u===IPv4===0x501b4d60181c8d01===0
t0===TCP===ESTABLISHED===130.126.138.124===54510===173.222.4.10===443
lsofMac===1430858964===1429669761===com.apple.WebKi===49972===absrdr===12u===IPv4===0x501b4d6018194431===0
t0===TCP===ESTABLISHED===130.126.138.124===54516===66.235.135.144===443
lsofMac===1430858964===1429669761===com.apple.WebKi===49972===absrdr===13u===IPv4===0x501b4d6018194431===0
t0===TCP===ESTABLISHED===130.126.138.124===54516===66.235.135.144===443
lsofMac===1430858964===1429669761===com.apple.WebKi===49972===absrdr===14u===IPv4===0x501b4d6017860771===0
```

Problems:

- Large volumes
- Difficult to analyze

Network-Resources View Data

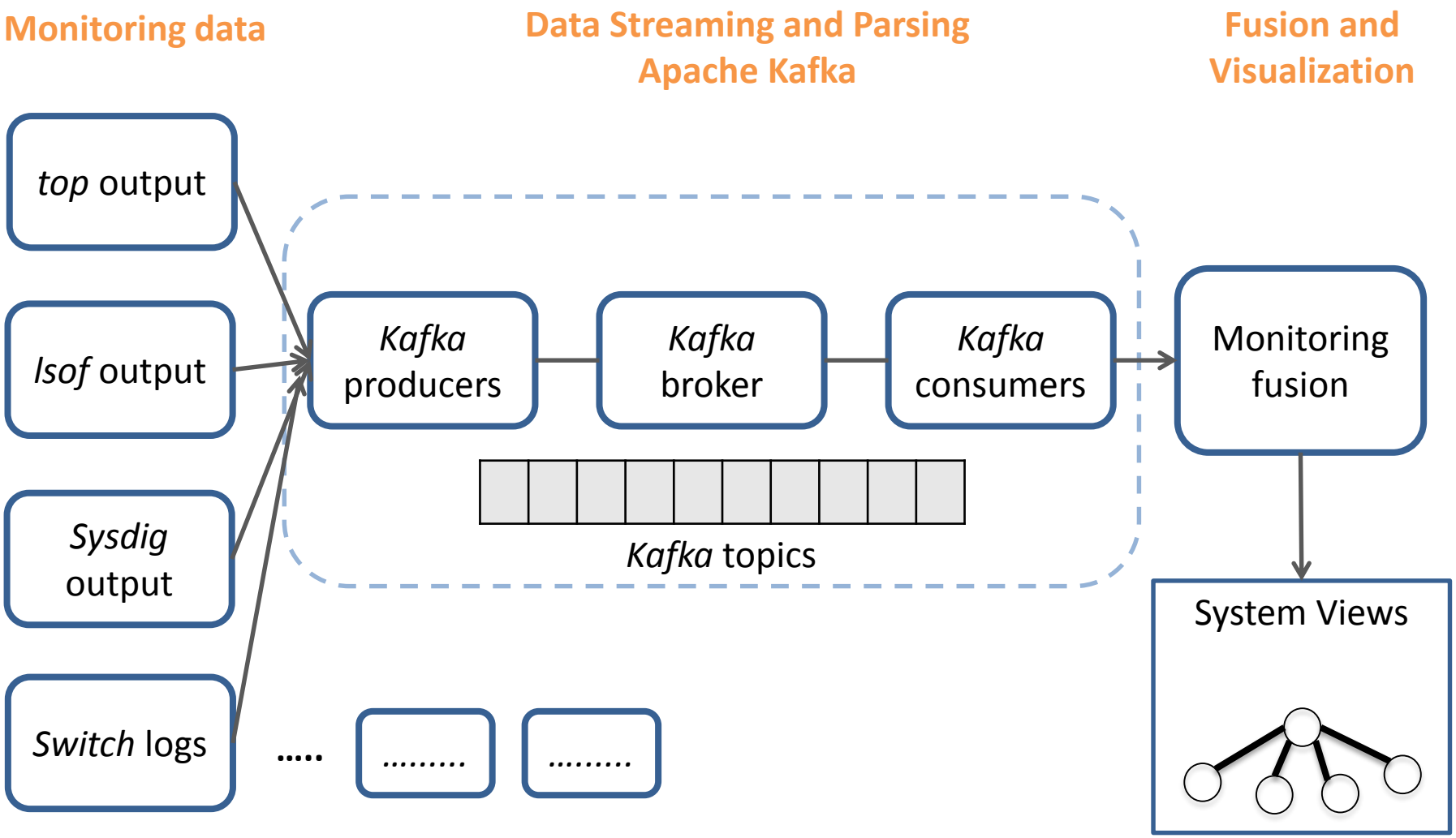
```
mongod ...    java    screen
Sending fused data to visualization server:
45477,127.0.0.1,127.0.0.1,java,0.0,299M+
Sending fused data to visualization server:
50298,130.126.138.124,74.125.202.97,com.apple.WebKit,0.0,23M+
Sending fused data to visualization server:
45477,127.0.0.1,127.0.0.1,java,0.0,299M+
Sending fused data to visualization server:
45474,127.0.0.1,127.0.0.1,java,0.0,61M+
Sending fused data to visualization server:
50298,130.126.138.124,74.125.202.97,com.apple.WebKit,0.0,23M+
Sending fused data to visualization server:
277,130.126.138.124,17.249.73.250,SpotlightNetHelp,0.0,8720K+
Sending fused data to visualization server:
45477,127.0.0.1,127.0.0.1,java,0.0,299M+
Sending fused data to visualization server:
45474,127.0.0.1,127.0.0.1,java,0.0,61M+
Sending fused data to visualization server:
277,130.126.138.124,17.249.73.250,SpotlightNetHelp,0.0,8704K+
Sending fused data to visualization server:
50298,130.126.138.124,74.125.202.97,com.apple.WebKit,0.0,23M+
Sending fused data to visualization server:
```

Fusion benefits

- Reduction in volume
- Easy visualization

	Monitors	Number of records	size
Raw data	top	20,219	1.2 MB
	lsof	14,930	1.7 MB
Fused data	**top+lsof**	**254**	**0.06 MB**

Fusion System Architecture



Outline

1. Generating System Views
    1. Handling Big and heterogeneous data

2. **Issues**
    1. **Conflicting evidences**
    2. **Independence of data sources**
    3. **Ensure security of the procedure**

Confidence in Fused Information

- Hypothesis: it's harder for an intrusion to deflect the output of multiple independent monitors

- Our approach
  - Utilize multiple independent monitors to generate same view
  - Accept the data if consistent otherwise report an alarm

Independence of Monitors

- Following classes of monitors are considered independent
  - Deployed at different physical locations
    - Host v/s network
  - Working at different access levels
    - User v/s kernel

Conclusion

- Monitoring fusion allows us to generate system views
- System views: convert monitoring data to useful information
  - Provide more useful and concise information
  - Ability to drill-down as required
  - Improve efficiency of decision making processes
  - Sophisticated algorithms can be built on top
- Set of views → System Activity Monitor

Future Work

- Build a catalog of useful views
- Data collection from multiple hosts and network devices
- Streaming + offline analytics
- Ensuring security of the process

Previous Related Literature

Fusion techniques for information security

- Data to information
  - Aggregation, data reduction techniques [Valdes'01, Ma'08, Czejdo'12]
- Information to knowledge
  - Alert correlation, multiple classifier combination [Cuppens'00, Debar'01, Ning'02, Ning'04, Totel'04, Zhu'06, Li'07, Almgren'08, Li'10, Roschke'10, Zhou'11, Kumar'14]

Data-driven approaches in intrusion detection

- Data mining and machine learning techniques in intrusion detection [Lee'08, Lan'10, Zhang'12, Yen'13, Kiss'14, Miller'14]

Online clustering algorithms

- K-means based [Aggarwal'03, Ackermann'12]
- Density based [Cao'06, Chen'07, Ding'15]

Previous Related Literature

Distributed System Management and Monitoring

- Bro Network Security Monitor
- Nagios IT infrastructure monitoring
- Distributed system resource managers
  - Mesos
  - YARN
  - OpenStack

Backup Slides

References

- [Valdes'01] Valdes, A., & Skinner, K. (2001, January). Probabilistic alert correlation. In *Recent advances in intrusion detection* (pp. 54-68). Springer Berlin Heidelberg.
- [Ma'08] Ma, W., Tran, D., & Sharma, D. (2008, June). A study on the feature selection of network traffic for intrusion detection purpose. In *Intelligence and Security Informatics, 2008. ISI 2008. IEEE International Conference on* (pp. 245-247). IEEE.
- [Czedjo'12] Czejdo, B. D., Ferragut, E. M., Goodall, J. R., & Laska, J. (2012). Network intrusion detection and visualization using aggregations in a cyber security data warehouse. *Int'l J. of Communications, Network and System Sciences*, *5*(09), 593.

References

- [Cuppens'00] Cuppens, F., & Ortalo, R. (2000). LAMBDA: A Language to Model a Database for Detection of Attacks. *Recent Advances in Intrusion Detection*, *1907*, 197–216.
- [Debar'01] Debar, H., & Wespi, A. (2001). Aggregation and correlation of intrusion-detection alerts. *Recent Advances in Intrusion Detection*, 85–103.
- [Ning'02] Ning, P., Cui, Y., & Reeves, D. S. (2002, November). Constructing attack scenarios through correlation of intrusion alerts. In *Proceedings of the 9th ACM conference on Computer and communications security* (pp. 245-254). ACM.
- [Ning'04] Ning, P., Xu, D., Healey, C. G., & Amant, R. S. (2004, February). Building Attack Scenarios through Integration of Complementary Alert Correlation Method. In *NDSS* (Vol. 4, pp. 97-111).
- [Totel'04] Totel, E., Vivinis, B., & Mé, L. (2004, January). A language driven intrusion detection system for event and alert correlation. In *Proceedings at the 19th IFIP International Information Security Conference, Kluwer Academic, Toulouse* (pp. 209-224).
- [Zhu'06] Zhu, B., & Ghorbani, A. A. (2005). *Alert correlation for extracting attack strategies* (Doctoral dissertation, University of New Brunswick, Faculty of Computer Science).
- [Li'07] Li, J., Lim, D. Y., & Sollins, K. R. (2007, August). Dependency-based Distributed Intrusion Detection. In *DETER*.

References

- [Almgren'08] Almgren, M., Lindqvist, U., & Jonsson, E. (2008, January). A multi-sensor model to improve automated attack detection. In *Recent Advances in Intrusion Detection* (pp. 291-310). Springer Berlin Heidelberg.
- [Li'10] Li, W., & Tian, S. (2010). An ontology-based intrusion alerts correlation system.*Expert Systems with Applications*, *37*(10), 7138-7146.
- [Roschke'10] Roschke, S., Cheng, F., & Meinel, C. (2010, September). A flexible and efficient alert correlation platform for distributed ids. In *Network and System Security (NSS), 2010 4th international conference on* (pp. 24-31). IEEE.
- [Zhou'11] Zhou, C. V., Leckie, C., & Karunasekera, S. (2010). A survey of coordinated attacks and collaborative intrusion detection. *Computers & Security*, *29*(1), 124-140.
- [Kumar'14] Kumar, C. A. (2014, November). Intrusion Detection Model Using fusion of PCA and optimized SVM. In *Contemporary Computing and Informatics (IC3I), 2014 International Conference on* (pp. 879-884). IEEE.

References

- [Aggarwal'03] Aggarwal, C. C., Han, J., Wang, J., & Yu, P. S. (2003, September). A framework for clustering evolving data streams. In *Proceedings of the 29th international conference on Very large data bases-Volume 29* (pp. 81-92). VLDB Endowment.
- [Cao'06] Cao, F., Ester, M., Qian, W., & Zhou, A. (2006, April). Density-Based Clustering over an Evolving Data Stream with Noise. In *SDM* (Vol. 6, pp. 328-339).
- [Chen'07] Chen, Y., & Tu, L. (2007, August). Density-based clustering for real-time stream data. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 133-142). ACM.
- [Ackermann'12] Ackermann, M. R., Märtens, M., Raupach, C., Swierkot, K., Lammersen, C., & Sohler, C. (2012). StreamKM++: A clustering algorithm for data streams.*Journal of Experimental Algorithmics (JEA)*, *17*, 2-4.
- [Ding'15] Ding, R., Wang, Q., Dang, Y., Fu, Q., Zhang, H., & Zhang, D. (2015). YADING: fast clustering of large-scale time series data. *Proceedings of the VLDB Endowment*, *8*(5), 473-484.

References

- [Lee'08] Lee, K., Kim, J., Kwon, K. H., Han, Y., & Kim, S. (2008). DDoS attack detection method using cluster analysis. *Expert Systems with Applications*, *34*(3), 1659-1665.
- [Lan'10] Lan, F., Chunlei, W., & Guoqing, M. (2010, April). A framework for network security situation awareness based on knowledge discovery. In *Computer Engineering and Technology (ICCET), 2010 2nd international conference on*(Vol. 1, pp. V1-226). IEEE.
- [Zhang'12] Zhang, J., Berthier, R., Rhee, W., Bailey, M., Pal, P., Jahanian, F., & Sanders, W. H. (2012, June). Safeguarding academic accounts and resources with the university credential abuse auditing system. In *Dependable Systems and Networks (DSN), 2012 42nd Annual IEEE/IFIP International Conference on* (pp. 1-8). IEEE.
- [Yen'13] Yen, T. F., Oprea, A., Onarlioglu, K., Leetham, T., Robertson, W., Juels, A., & Kirda, E. (2013, December). Beehive: Large-scale log analysis for detecting suspicious activity in enterprise networks. In *Proceedings of the 29th Annual Computer Security Applications Conference* (pp. 199-208). ACM.
- [Kiss'14] Kiss, I., Genge, B., Haller, P., & Sebestyen, G. (2014, September). Data clustering-based anomaly detection in industrial control systems. In *Intelligent Computer Communication and Processing (ICCP), 2014 IEEE International Conference on* (pp. 275-281). IEEE.
- [Miller'14] Miller, Z., Dickinson, B., Deitrick, W., Hu, W., & Wang, A. H. (2014). Twitter spammer detection using data stream clustering. *Information Sciences*, *260*, 64-73.