



Visualization Techniques for Intrusion Detection

William Wright and Peter Clarke

Point of Contact: William Wright Oculus Info Inc. 572 King Street West, Suite 200 Toronto, Ontario M5V 1M3 CANADA

bill.wright@oculusinfo.com

INTRODUCTION

This paper reports on the experiences of using interactive animated 2D and 3D graphics in an Intrusion Detection (ID) Analysts Workbench prototype. Visualization techniques allow people to see and comprehend large amounts of complex data. Graphics are used to assist with the ID investigation and reporting process by helping the analyst identify significant incidents and reduce false conditions (positives, negatives and alarms). Visualization is then used in reporting incidents to a broader senior level audience. Complex patterns are clearly displayed over time in an easy to understand and compelling manner. Initial evaluations of the prototype have been positive, and a second development stage has been initiated.

ID ISSUES

Large numbers of events are generated by network intrusion detection sensors; however not all these events are malicious in nature, not all malicious events are applicable to a given network environment and, perhaps of even more concern, certain malicious events can be missed.

There are several emerging trends in enterprise networking that are making traditional signature based intrusion detection more challenging. The increase use of very high-speed lines and more prevalent use of encryption technology are a challenge for the intrusion detection community. As the data collected becomes larger in volume, or the increasing dependence on traffic pattern anomaly detection as a workaround for payload encryption becomes more widespread, the amount of data the analyst must cope with increases.

Through the use of various types of detection tools and techniques, including signature based network intrusion detection, anomaly based network intrusion detection, and full packet capture, a better picture can be formed. The analyst is able to fuse this data and gain a more comprehensive insight into what is truly of malicious nature.

The massive amounts of data involved in this type of thorough multi-source analysis make it infeasible for most organizations. The significance of the events contained within the data can often only be determined by scanning the huge amounts of data looking for subtle and sometimes unexpected patterns and correlations.

This investigative process is required in order to place the events around an alarm in context and to assess if further action is required, but the process is labor intensive. Fused logs for a short period can easily contain tens to hundreds of thousands of records. Tools are needed to help accomplish this investigative task in less time.

Paper presented at the RTO IST Workshop on "Massive Military Data Fusion and Visualisation: Users Talk with Developers", held in Halden, Norway, 10-13 September 2002, and published in RTO-MP-105.

	Report Documentation PageForm Approved OMB No. 0704-0188					
maintaining the data needed, and c including suggestions for reducing	lection of information is estimated to completing and reviewing the collect this burden, to Washington Headqu uld be aware that notwithstanding ar DMB control number.	ion of information. Send comments arters Services, Directorate for Infor	regarding this burden estimate mation Operations and Reports	or any other aspect of the s, 1215 Jefferson Davis	is collection of information, Highway, Suite 1204, Arlington	
1. REPORT DATE 00 APR 2004		2. REPORT TYPE N/A		3. DATES COVE	RED	
4. TITLE AND SUBTITLE				5a. CONTRACT	NUMBER	
Visualization Tech	niques for Intrusion	Detection		5b. GRANT NUM	/BER	
				5c. PROGRAM E	LEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NU	JMBER	
				5e. TASK NUMB	BER	
				5f. WORK UNIT	NUMBER	
7. PERFORMING ORGANI Point of Contact: V Suite 200 Toronto,	8. PERFORMING ORGANIZATION REPORT NUMBER					
9. SPONSORING/MONITO	RING AGENCY NAME(S) A	AND ADDRESS(ES)		10. SPONSOR/M	ONITOR'S ACRONYM(S)	
				11. SPONSOR/M NUMBER(S)	ONITOR'S REPORT	
12. DISTRIBUTION/AVAII Approved for publ	LABILITY STATEMENT ic release, distributi	on unlimited				
	otes 65, RTO-MP-105 M riginal document co	-		ualization: U	sers Talk with	
14. ABSTRACT						
15. SUBJECT TERMS						
16. SECURITY CLASSIFIC	CATION OF:	17. LIMITATION OF	18. NUMBER	19a. NAME OF		
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified	ABSTRACT UU	OF PAGES 26	RESPONSIBLE PERSON	

Standard Form 298 (Rev. 8-98	3
Prescribed by ANSI Std Z39-1	ŝ



Another issue is that network ID sensors are not always effective for detecting new exploits or for activities that span many weeks and / or multiple network systems. To detect and investigate these types of activities requires the analyst to review extremely large amounts of packet level data.

A related problem is in reporting the attacks and the nature of those attacks to senior managers. This is important in order to raise awareness and provide an understanding of the need for information technology security in industry and government. Without this senior level understanding and support, obtaining security funding can be challenging. The output of most ID processes can be cryptic, and inaccessible to non-experts.

SOLUTIONS

Two graphical consoles have been built to evaluate the usefulness of visualizing intrusion data. Figure 1 shows the Intrusion Detection Analysts Workbench. Up to 2,000,000 event records or more can be displayed and analyzed in multiple concurrent dynamic charts. Each event record includes fields such as source and destination IP, port ID, alarm code, date, time. The charts are scaleable so that, for example, a bar chart showing number of events by destination IPs can easily display ten's of thousands of IP addresses. The charts are also linked. Selecting events in one chart will highlight those events in all the other charts. So for example, selecting events associated with one type of alarm will cross reference those events in the source IP bar chart, and the destination IP bar chart. The analyst workbench is used to investigate, isolate and prioritize events. It was evaluated in a side-by-side test with existing methods and proved to be a significantly faster method. The workbench makes use of the commercial off-the-shelf Advizor product.

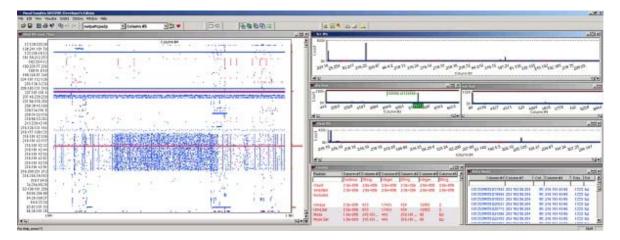


Figure 1: Intrusion Detection Workbench with 2M TCP and UDP Records.

The Analysts Workbench graphical tool can concurrently display the raw packet or alarm data as well as output from analytical tools that, for example, filter events or compute statistical metrics.

Figure 2 shows the Animated Incident Reporting component. It is used to report intrusion activity to senior management, and is designed to show the significance and nature of the events without overwhelming the viewer. The objective is to clearly see who did what to whom and when. A number of interactions are supported including filtering and an adjustable playback speed. This component was evaluated in a series of presentations to senior levels of government and industry.



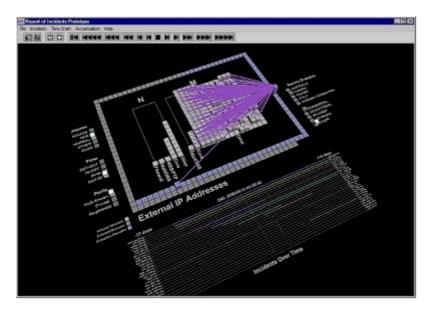


Figure 2: Animated Incident Reporting.

FUTURE DEVELOPMENT

Future work involves two separate but related streams:

The first is the expansion and integration of the two visualization tools to create a seamless intrusion detection visualization workflow environment. Given that intrusion detection analysis is often only part of a systems administration function, time is a consideration. The more effectively the visualization tools can be adapted to fit, and enhance, the human decision making process (orient, observe, decide and act), the more incidents can be effectively assessed and escalated or discarded in a shorter time period.

The second is work on migrating the tools towards an anomaly detection capability through the use of raw network data along with the fused intrusion detection alarms to gain a more comprehensive view into the network.

CONCLUSION

People excel in detecting patterns and identifying relationships when data is presented visually. Extremely large amounts of data can be viewed and compared. This is a useful ability for the ID analyst.

Experience with the visualization methods used in this work has lead to observations and recommendations for developing new methods.

Initial evaluations of the prototype have been positive, and a second development stage has been initiated. The objectives of this second stage will be discussed in the paper.



SYMPOSIA DISCUSSION – PAPER NO: 15

Author's Name:

Mr. William Wright, Oculus Info Inc, Canada Presented by Ms. Pascale Proulx, Oculus Info Inc, Canada

Question:

What type of methodology was applied in the design of the visualisation?

Author's Response:

User consultations.

Comment:

The system is ten times faster than the previous system that did not use any visual display at all.

Comment:

These display techniques would be useful for all statistical data such as traffic jams.

Question:

Is it possible to transfer the knowledge of the operator into rules for the system to automate the process?

Author's Response:

It may be possible to recognize the patterns, but it is important to investigate more to see what is generating the pattern. For example, a pattern that initially appears dangerous may only be a virus definition update.

Question:

In advance of developing this system, was there consideration of mathematical methods that maybe amenable to clustering and statistical analysis?

Author's Response:

There is a trade off between the math and the visual analysis, as well as between implementation time of the math and algorithms and human interaction. There is research going on in this area right now.

Comment:

Change detection might be important to an analyst.

Question:

There are clustering techniques that could be applied, but require a large amount of training data to make the systems work well. How much data is available for training?

Author's Response:

There currently are not many full data sets. There is a conference in LA that puts hackers against professions, but that produces a data set that is not necessarily typical. Also, networks are dynamic, so constant retraining is required.



Increasing Visibility into Intrusion Detection

Visualization Techniques for Intrusion Detection

William Wright, Sr. Partner, Oculus, Canada <u>bill.wright@oculusinfo.com</u>

> and Peter Clarke, DND, Canada

Presented by Pascale Proulx, Visualization Consultant, Oculus pascale.proulx@oculusinfo.com

Paper first given at: <u>Workshop on Statistical and Machine Learning Techniques in</u> <u>Computer Intrusion Detection</u>, Johns Hopkins University, June 11-13, 2002.



2

What is Covered?

- Intrusion detection issues
- Using visualization as a solution
- Current visualization tools developed
- Future development of visualization in intrusion detection

Intrusion Detection Issues

- Large amounts of IDS (Intrusion Detection Sensor) data
 - 1 gigabyte of information will fill a pickup truck with printed paper
- Bad signal/noise ratio on most un-tuned IDS
 - Worse than TV filled with "snow"
- If alarms are removed, harmful events may slip through unnoticed
- Event Correlation (IDS, routers, firewalls)
 - Very important to gaining a complete picture, but makes data handling even more difficult



Intrusion Detection Issues

- Reporting incidents to senior management or other non-experts
- The problems are getting worse as technology progresses and network speeds (i.e. bandwidth) increase

5

Visualization as a Solution

- Visualization allows people to see and comprehend large amounts of complex data in a short period of time
- Helps the analyst to identify significant incidents and reduce time wasted with false positives
- Report incidents to a broader, non-expert audience
- Ability to cue the analyst through the use of colour, shape, patterns, or motion

6

Visualization Tool Development

- Two graphical applications have been built for evaluation
 - Intrusion Detection Analyst Workbench
 - Animated Incident Explanation Engine
- Each displays data visually, but currently have two separate audiences

Intrusion Detection Analyst Workbench

- More than 2 million events can be displayed and analyzed in multiple concurrent dynamic charts
- Each chart is linked, allowing the analyst to select something in one chart, and it will highlight the relevant details in the other charts
- High performance interactive analysis possible for 2M records on high end Windows / Intel machine 2 GHz, 1 GB RAM and good graphics card.

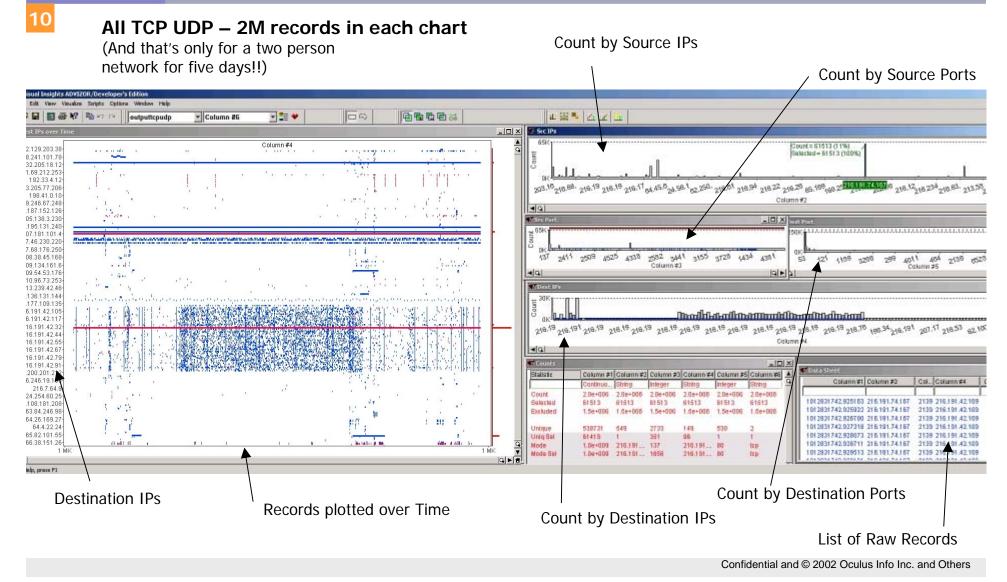
Intrusion Detection Analyst Workbench

- Assists in isolating, investigating and prioritizing events
- Evaluated side-by-side with existing methods and proved to be significantly faster and easier
- Run by commercial off-the-shelf Advizor™ product

Log Data Fields

- Date
- Time
- Direction out-in, in-out, out-out, in-in
- Alarm Code (alarms and/or normal traffic)
- Alarm Name
- Source IP
- Destination IP
- Port ID Source
- Port ID Destination
- Port Name ~10 well known, 65,000 possible
- Sensor ID

oculus **Intrusion Detection Analysts Workbench** Layout consists of multiple charts. Each chart has all the records.



15-10

	culu	S			Lin	ked	Cha	rtino	r			
										25500	ciated with	this
											are highligh	
10 553	N 21 22	an l							occur in the		0 0	neu
Sic IPs										/	graphor	-
							· ·					
65K							Count = 61: Selected =	61.51.3 (10.096)				
Cau	<u>10</u>		n				Constants.				- 27	
ال المالية		<u>بالم</u>	والمساليات		- 1.		in the second	1223		-		
203,10,210.	38. 218.19 218	18 218.57	4.45 P 24.56	3. 1 12.258.	218.01 216	394 218 22 2	196,20 es.40	198 Z 210.18	14 10 218.	21823	14 218.83 213	1218,11
							lumn #2			-		2
10												
Sre Part						느므즈	hast Past					
457 244	4 25109 46Z	5 ₄₃₀₈	<u>7582</u> 344 Column≢3	1 3155	3T28 44	1854 AZ	0K 53 47	d 1188 3	599 259 M	gt i A Salumnia	104 2138 65 #5	<u>, z</u> si 1
Const De	4 2009 A62	5 ₄₃₈ 8		i ₃₁₅₅	5TZ8 44	St. 194064 (C	100	d 1188 3	259 259 M	AN A Salumn a	0 ⁴ 21 ³⁰ න් #5	zs 4
			Column #1	madhta	a a a filma	5 ► 100000000000000000000000000000000000	191 100-00100 1,19 216,19 10m #4	Maaad			194 2198 53	
	, ¹⁹¹ 2 ^{18,19} 2	18.19 216.1	Column#1	Denet[[1]]e 2 ^{1/67,19} 2 ^{1/8}	1601100 3,19 ₂ (6),15	اعا⊾ 1110000000000000000000000000000000000	191 100-00100 5,49 216,19 Num #4	Minana 2 ^{(6,19} 2 ^{(6,1}			15 	
	1 ⁴⁹¹ 2 ^{18,59} 2 Column #1	18.19 216.1	Column#1	Denet[[1]]e 2 ^{1/67,19} 2 ^{1/8}	1 1-7111111 5. ¹⁶ 2 ^{16,15} Calumn #5	ع ا- 11 2 روت ⁽¹⁹ 2 رو Col ا <u>حال</u>	191 100-00100 5,49 216,19 Num #4	1:1110-11:111 2 ^{16,19} 218, ¹¹ Shee		1 ¹ 2017	15 	100 216
	1 ⁴⁹¹ 2 ^{49,49} 2 Column #1	(5. ^{1,5} ₂ \6.1 Column≠2	Column #3	Donallala t ^{ya,19} 218 Column#4	1601100 3,19 ₂ (6),15	ع ► 1 1 1 2 ^{16,19} 2 ¹⁶ Col 1 2 ¹⁰	191 1000000000 3,49 216,19 Num #4	1:1110-11:111 2 ^{16,19} 218, ¹¹ Shee	L	1 ¹ 2017	*5 ⁴¹ 218,53 62.1	100 218
Count Statistic	Column #1 Continuo 2 De+000 61 51 3	(6. ^{1,6} 216. ¹ Column#2 Sting 2.0e+005 61512	Column #3	Column #4 String 2.0e+005 61513	Column #5	(11) 1215,19 219 Cal Cal Cal Cal Cal Cal Cal Cal	1)11 1)11 1)19 1,19 216,19 Nmn #4	Hington 2 ^{16,19} 2 ^{16,1} 216,5 216,5 21742,9251 63	Calumn #2	¹¹ ₂ 01; Сон. 2139	≠5 11 _{216,5} 3 ₆₂ Column≠4 216,191,42,109	100 218
Count	Column #1 Continue 2 De+000	(6. ^{1,6} 216.1 Column#2 Sting 2.0e+005	Column #3	Column #4 Sting 2.0e+005 61513	0	(13) • 13) • 1 21 ^{5,19} 21 ⁶ Col 215,19 216 Col 200 200 000	10128 10128	Hindrad 2 ^{10,19} 2 ^{10,11} Shee: Column#1 31742 5251 63 31742 5251 63	³ 1,55.3 ⁴ ,216,15 Column ≠2 216.191.74.167 218.191.74.167	¹ ₂ сл. с.н. 2139 2139	≠5 ⁴¹ 216,53 62. ¹ Column ≠4 216,191,42,109 216,191,42,109	100 218 Colu.
Count Statistic	Column #1 Column #1 Continue 2 De+006 Bt 51 3 1.5e+006	(6. ^{1,6} 216. ¹ Column#2 Sting 2.0e+005 61512	Column #3 (Column #3 (Column #3 (Neper 2.0e+006 £1513 1.5e+006	Column #4 String 2.0e+005 61513	Column #5	14 + 141100000101 1 216,19 216 Cal Cal 216,19 216 Cal Cal 2161 200+006 61513 1.5+006	10128 10128 10128 10128 10128 10128	Hindind 2 ^{16,19} 218, ¹¹ 216,19 216,19 21742,9261,63 21742,9261,63 21742,9261,63 21742,9261,63 21742,9261,63	Calumn #2	¹ 201. 201 2139 2139 2139 2139 2139	≠5 11 _{216,5} 3 ₆₂ Column≠4 216,191,42,109	100 218 Colu.
Statistic Count Statistic Unique	Column #1 Column #1 Continue 2 De+006 Bt 51 3 1.5e+006	(6,1 ⁶ 216,1 Column#2 String 2.0e+005 61513 1.5e+005	Column #3 (Column #3 (Column #3 (nieper 2. De+006 £1 513 1. 5e+006 £1 513 1. 5e+006 £1 513 1. 5e+006	Column #4 String 2.0e+005 61513 1.5e+005	Colume #5 (Colume #5 (nieger 2.De+006 61513 1.Se+006	(11) 1215,19 219 Cal Cal Cal Cal Cal Cal Cal Cal	10128 10128 10128 10128 10128 10128 10128	50000000000000000000000000000000000000	Column #2 216.191.74.167 216.191.74.167 216.191.74.167 216.191.74.167 216.191.74.167 216.191.74.167	^A 201. 2139 2139 2139 2139 2139 2139 2139 2139	4 ¹ 216,5 ³ 62, ¹ Column #4 216,191,42,109 216,191,42,109 216,191,42,109 216,191,42,109 216,191,42,109 216,191,42,109	
Count Statistic Unique	Column #1 Continue 2 De+000 Bt 51 3 1.0e+000 Colle721 Bt 410 1.0e+009	Column#2 String 2.0e+005 61513 1.5e+005 649	Column#3 (Column#3) (Column#3) (nieper 2.0e+000 E1513 1.5e+000 21739 251 1.37	Column #4 String 2.0e+005 61513 1.5e+005	Columer #5 integer 2.De+006 81513 1.5e+006 1 80	14 + 141100000101 1 216,19 216 Cal Cal 216,19 216 Cal Cal 2161 200+006 61513 1.5+006	1012 1012 1012 1012 1012 1012 1012 1012	Sheet Sheet Column#1 21742.925163 21742.925163 21742.925922 21742.925923 21742.925973 21742.925973 21742.925973	Calumn #2 216.191.74.167 216.191.74.167 216.191.74.167 216.191.74.167	Col. 2139 2139 2139 2139 2139 2139 2139 2139	⁴¹ 216,53 62. ¹ Column ≠4 216,191,42,109 216,191,42,109 216,191,42,109 216,191,42,109	100 218 100 218 100 218 100 218

Linked Charting

2 🖪 🍜 K	🧟 🖄 🗠 🖓	estputtepudp	💌 Column #6	<u>-</u>		🖣 🖥 🖷 🚟		_]] ≣ ∰
t DPs over Tim	e							F Sec IPs
02.186.160				Column≠4				_ 65(
6.34.1Z.12		1. A.						
09.200.244		4						2 I I I
12.206.194								المقيرة (10
2.74.101.9								203,10 2101
12.251.20								
91.42.102			"\$46 TE - 11 KE		149 W 14	1122日 日本人間		• Q
91.42.106			없는 것 같은 말을 가 있다.			나는 말 잘 못 했네?		Src Pert
91.42.110 91.42.114	1.1.15			H.S. C.A. N	(Hu MD Ha	te sint sint		GOK
91.42.119			쓰장기 가난다?	andro de la	신 옷을 빼내었다.			
91.42.122			6 K. L. H. N.		- 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1			481 ⁵⁴
91.42.126 191.42.33		그 옷에 앉아요.		ti je da se	1917 - M ARI S			
191.42.38		t de block	TME EPZ EDALG	de la factoria de la composición de la	이 문제 않는 것이 같은	tta Marka Mark		<u> 9 </u>
191.42.42			화는 17 귀엽이다.			나는 지원에 관심하지 않는		Dest IPs
191,42,46			승규는 문제가 같아?	김 국가 같은	i de Qérier			30Kg
191.42.54			12. R. M. M. M. M. M.		- 19 8			
191.42.58			지난 이 이 지하는		- 14 - 100 - 10		· · · · · · · · · · · · · · · · · · ·	ت _{الا} لی
191.42.62			104 I.J. (1914)		e të e të ka kë k			218,18 24
191,42,70				H. I. Scherf	1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 -			the the
191.42.74			20년 17 - 신유민의		114 B U ARA			<[G]
191.42.78 191.42.82		A. 14 II. A. 194	보안 다. 너 같은		1 N N N			Counts
191.42.86			승규는 비가 있		同時の間が日本	승규는 김 씨는 것 같아?		Statistic
91.42.90			74 I J I I.	김희 동안 집 집	이 같은 행사 같은 것			
191.42.94	국가 가 같		영제 바라다		에 걸려 있는 것이 같이 많이 많이 많이 했다.			Count
05.212.89			7 - 0 - 1 - 1 - 0 - 1 - 1 - 1 - 1 - 1 - 1					Selected
8.230.178				1			· · · · · · · · · · · · · · · · · · ·	Excluded
78.108.76				-				(Interes
0.129.146						1 A A A A A A A A A A A A A A A A A A A		Unique Uniq Sei
із 193 жей. 1 ЙК							8 S. S. H. H.	Node

For Help, press P3

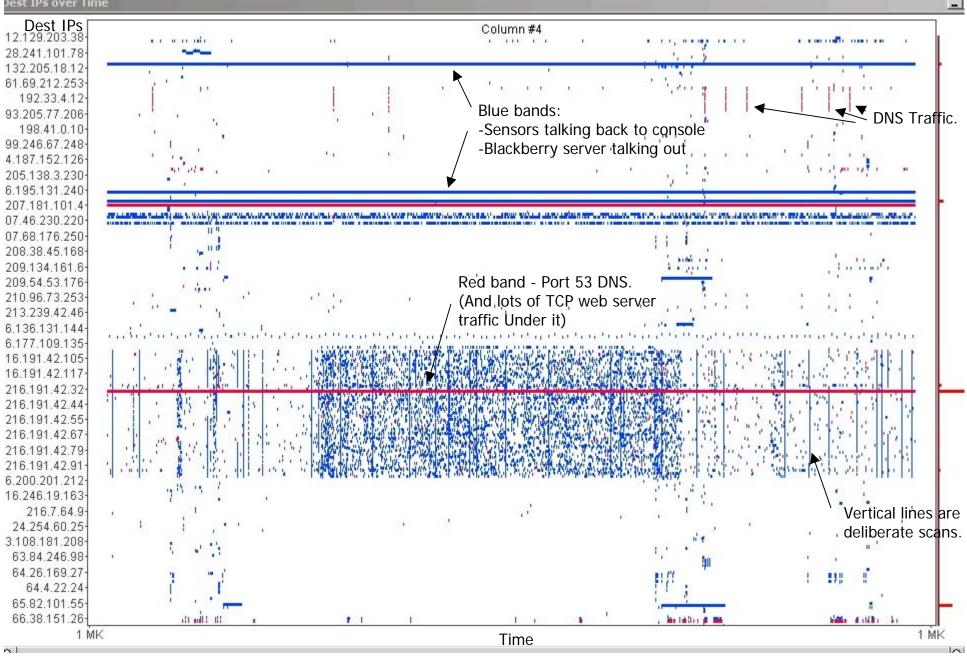
oculus

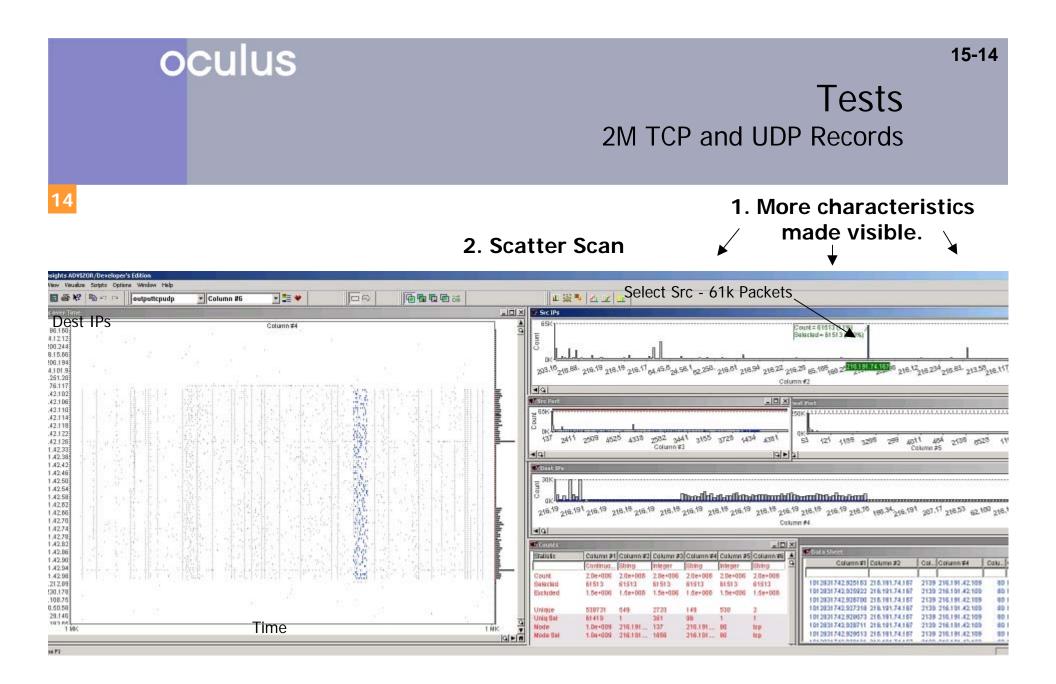
15-12

15-13 Analysis - All TCP UDP – 2M Records

TCP is Blue and UDP is Red. Five days. Two people.

est IPs over Time

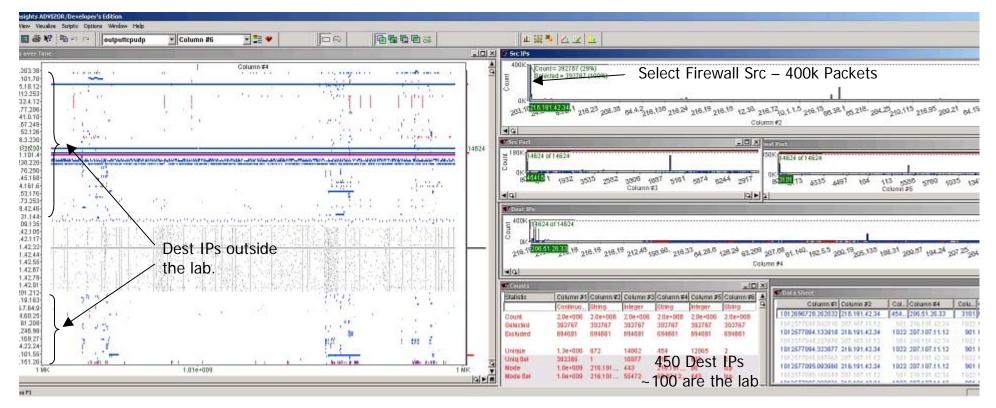




15

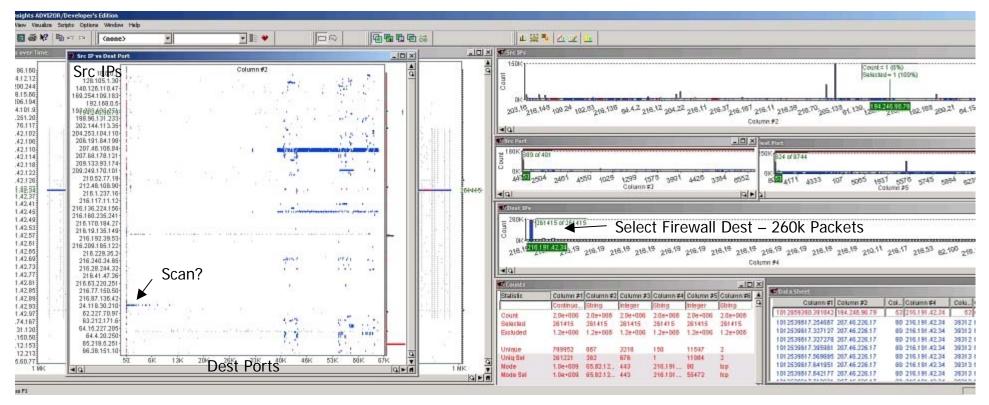
Tests 2M TCP and UDP Records

Traffic Outbound – Firewall Src



Tests 2M TCP and UDP Records

Low Port Scan?



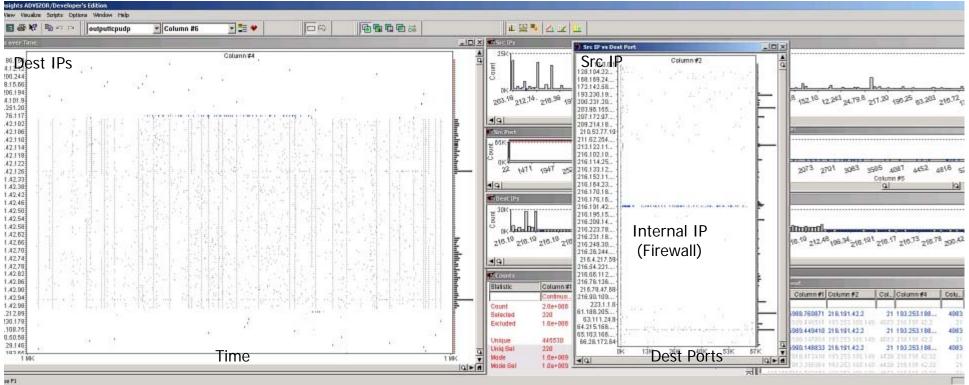
15-16



Tests 2M TCP and UDP Records

17

Slow Scan? one Src, many ports, over time.



On the weekend

Or Microsoft virus definition file update.



18

Animated Incident Explanation Engine

- Designed to show the significance and nature of the events without overwhelming the viewer
- Easy to see who did what to whom and when
- Excellent for explaining concepts to non-experts
- Analyst workbench identifies and isolates incidents. When an incident is isolated, all the records associated with it are written out and saved. Over time, a set of verified assessed incidents is created.
- The explanation engine works on this set of incidents.

Animated Incident Explanation Engine

IP Half Scan is shown.

Accessition Help **14 3444 344 34 3** External IP Addresse Timeline of Incidents -one line per incident HE CARE TH

Usage – show all records, show all records associated with one incident, animate over time all records, or animate records for one or more incidents. Playback speed is from fast fwd to very slow.

15-19

Internal IPs -grouped by function

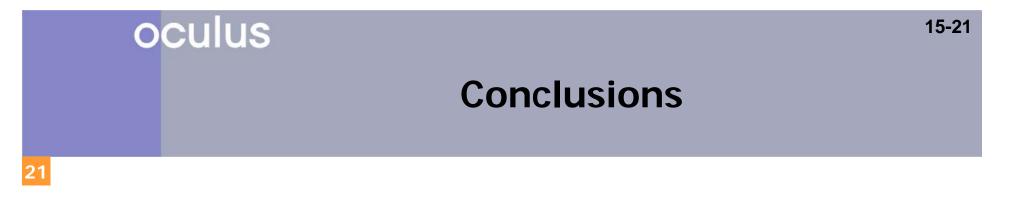
External IPs - sorted by frequency - and/or grouped by watchlist order



20

Future Developments

- Expansion and integration of the two current tools
- Anomaly detection capability through the use of network traffic data along with fused IDS alarms
- Integrated time based comparisons



- Visualization has proved to be an effective analyst's tool
- Massive amounts of information are easily understood by non-experts
- More development and research needed



Questions?