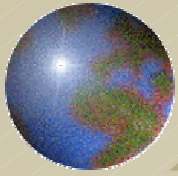


Data Mining NetFlow

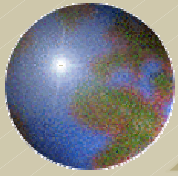
So What's Next?

Mark E Kane
FloCon 2005
20 September 05



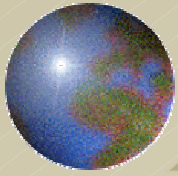
Objectives

- ❖ Data Mining, very briefly
- ❖ Frequency Patterns
- ❖ Discoveries
- ❖ Realizations
- ❖ Changes Made



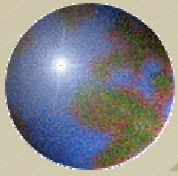
Data Mining

Data Mining – automated extraction of previously unknown data that is interesting and potentially useful.



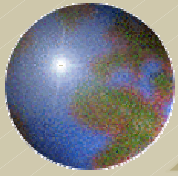
Cost of Participating in Data Mining

| Reality | Result of Data Mining | Example Analyst Hours | Example Investigator Hours | Example SysAdmin Hours | Result |
|---------|-----------------------|-----------------------|----------------------------|------------------------|---|
| YES | YES | 10 | 10 | 10 | Crime Prevented / Prosecuted |
| NO | NO | 0 | 0 | 0 | - |
| YES | NO | ∞ | ∞ | ∞ | Time Lost to Investigate and Clean Up After Crime |
| NO | YES | 10 | 10 | 10 | Red Haring |



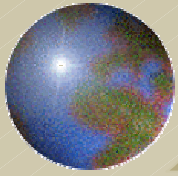
Complexity of Mining NetFlow

- ✚ Shear Volume
- ✚ Complex Protocol Analysis
- ✚ Ambiguous Interpretations
- ✚ Very Smart Adversaries



Common Investigator Issues

- ❖ Undermanned and overworked
- ❖ Varied knowledge base
- ❖ Does not own networks
- ❖ No direct reporting structure



Data Mining Techniques

Primary Techniques

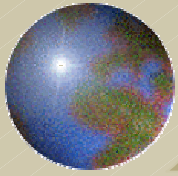
- Rule and Tree Induction
- Characterization
- **Classification**
- Regression
- Association
- **Clustering**

Other Techniques

- Dependency Modeling
- **Change Detection**
- Trend Analysis
- Deviation Detection
- **Link Analysis**
- Pattern Analysis
- Spatiotemporal Data Mining
- Mining Path Traversal Patterns
- **Mining Sequential/Frequent Patterns**

Uncertain Reasoning Techniques

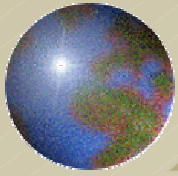
- Fuzzy Logic
- Neural Networks
- Bayesian Networks
- Genetic Algorithms
- Rough Set Theory



Frequency Patterns

Mining Frequent Patterns in Data Streams in Multiple Time Granularities (Giennella, Han, Pei, Yan, and Yu)

- ❑ Support Decision Making
- ❑ Past Less Significant than Present
- ❑ Record Reduction
- ❑ Time Tilted Windows



Interpreting Time-Tilted Windows

| | DAY | | | | | | | | |
|------------|-----------|----|----|----|----|---|---|---|---|
| Window | | 0 | | 1 | | 2 | | 3 | |
| Transition | | N | Y | N | Y | N | Y | N | Y |
| Size | | 1 | 1 | 2 | 2 | 4 | 4 | 8 | 8 |
| | Monday | 9 | | | | | | | |
| | Tuesday | 15 | 9 | | | | | | |
| | Wednesday | 6 | | 12 | | | | | |
| | Thursday | 6 | 6 | 12 | | | | | |
| | Friday | 12 | | 6 | 12 | | | | |
| | Saturday | 16 | 12 | 6 | 12 | | | | |
| | Sunday | 6 | | 14 | | 9 | | | |
| | Monday | 12 | 6 | 14 | | 9 | | | |
| | Tuesday | 15 | | 9 | 14 | 9 | | | |

Day 1: 9 events

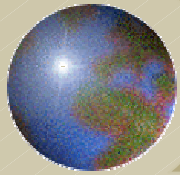
Day 2: 15 events (two buckets)

Day 3: 6 events (two buckets)

Day 4: 6 events (two buckets)

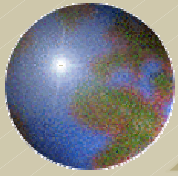
Day 5: 16 events (three buckets)

Day 6: 12 events (four buckets)



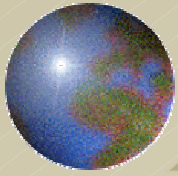
Presenting Frequency Patterns

| Address | Window Number | Byte Support | | | | | | | | | | Transaction Support | | | | | | | | | | Packet Support | | | | | | | | | | | |
|-----------|--------------------|--------------|------|------|------|------|------|------|------|------|-------|---------------------|------|------|------|------|------|------|------|------|------|----------------|------|------|------|---|------|-------|------|------|------|------|------|
| | Transition Ind | 0 | | 1 | | 2 | 3 | 4 | 5 | | 0 | | 1 | | 2 | 3 | 4 | 5 | | 0 | | 1 | | 2 | 3 | 4 | 5 | | | | | | |
| | Window Size (Days) | N | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N | Y | | | | |
| 5.228.76 | 11.00 | | | | | | | | | | 0.00 | | | | | | | | | | | | | | | | | 2.13 | | | | | |
| 162.12 | 6.96 | | | | | | | | | | 0.00 | | | | | | | | | | | | | | | | | 1.26 | | | | | |
| 160 | 4.09 | | | | | | | | | | 0.00 | | | | | | | | | | | | | | | | | 0.86 | | | | | |
| 3.16.11 | 3.06 | | | | | | | | | | 3.69 | | | | | | | | | | | | | | | | | 2.23 | | | | | |
| 5.32.21 | 2.62 | 0.63 | 0.75 | | | | | | | | 0.00 | 0.01 | 0.00 | | | | | | | | | | | | | | | 0.90 | 0.57 | 0.62 | | | |
| 5.238.67 | 2.42 | | | | | | | | | | 0.51 | | | | | | | | | | | | | | | | | 0.89 | | | | | |
| 5.235.66 | 2.17 | 0.17 | | | | | | | | | 67.67 | 4.00 | | | | | | | | | | | | | | | | 16.10 | 0.72 | | | | |
| 46.43 | 2.16 | | | | | | | | | | 0.00 | | | | | | | | | | | | | | | | | 0.95 | | | | | |
| 73.1 | 2.00 | 0.00 | 0.00 | 0.00 | | | | | | | 0.16 | 0.01 | 0.01 | 0.09 | | | | | | | | | | | | | | 0.49 | 0.00 | 0.02 | 0.01 | | |
| 5.48.212 | 1.86 | 0.57 | | | | | | | | | 0.22 | 0.22 | | | | | | | | | | | | | | | | 0.49 | 0.33 | | | | |
| 3.252.103 | 1.76 | | | | | | | | | | 6.99 | | | | | | | | | | | | | | | | | 2.99 | | | | | |
| 2.168.35 | 1.72 | 0.00 | 0.10 | | | | | | | | 0.00 | 0.00 | 0.00 | | | | | | | | | | | | | | | 0.43 | 0.00 | 0.03 | | | |
| 97.85 | 1.71 | | | | | | | | | | 0.00 | | | | | | | | | | | | | | | | | 0.48 | | | | | |
| 3.74.105 | 1.69 | | | | | | | | | | 0.14 | | | | | | | | | | | | | | | | | 0.50 | | | | | |
| 5.232.159 | 1.61 | | | | | | | | | | 0.53 | | | | | | | | | | | | | | | | | 0.62 | | | | | |
| 03.25 | 1.31 | | | | | | | | | | 0.06 | | | | | | | | | | | | | | | | | 0.24 | | | | | |
| 104.74 | 1.28 | | | | | | | | | | 0.00 | | | | | | | | | | | | | | | | | 0.10 | | | | | |
| 2.5 | 1.25 | | | | | | | | | | 0.00 | | | | | | | | | | | | | | | | | 0.41 | | | | | |
| 115.107 | 1.25 | | | | | | | | | | 0.08 | | | | | | | | | | | | | | | | | 0.32 | | | | | |
| 4.214.79 | 1.20 | | | | | | | | | | 0.00 | | | | | | | | | | | | | | | | | 0.26 | | | | | |
| 4.22.76 | 1.18 | | | | | | | | | | 0.97 | | | | | | | | | | | | | | | | | 0.58 | | | | | |
| 162.70 | 1.15 | | | | | | | | | | 0.31 | | | | | | | | | | | | | | | | | 0.42 | | | | | |
| 4.170.174 | 1.14 | | | | | | | | | | 0.46 | | | | | | | | | | | | | | | | | 0.44 | | | | | |
| 4.87.219 | 1.09 | | | | | | | | | | 0.34 | | | | | | | | | | | | | | | | | 0.39 | | | | | |
| 84.150 | 1.08 | | | | | | | | | | 0.10 | | | | | | | | | | | | | | | | | 0.38 | | | | | |
| 5.48.80 | 1.07 | | | | | | | | | | 0.26 | | | | | | | | | | | | | | | | | 0.30 | | | | | |
| 48.44 | 1.04 | 0.21 | 0.12 | | | | | | | | 0.00 | 0.00 | 0.00 | | | | | | | | | | | | | | | 0.19 | 0.07 | 0.03 | | | |
| 3.245.112 | 1.03 | 0.66 | | | | | | | | | 0.43 | 0.36 | | | | | | | | | | | | | | | | 0.67 | 0.41 | | | | |
| 60.148 | 1.01 | | | | | | | | | | 0.01 | | | | | | | | | | | | | | | | | 0.15 | | | | | |
| 3.245.160 | 1.00 | | | | | | | | | | 0.04 | | | | | | | | | | | | | | | | | 0.21 | | | | | |
| 152.150 | 0.09 | | 0.01 | | 0.00 | 0.00 | 0.01 | 0.01 | | | 0.01 | | 0.00 | | 0.00 | 0.00 | 0.00 | 0.00 | | | | | | | | | | 0.02 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 |
| 3.177.215 | 0.06 | 0.06 | 0.07 | | 0.02 | 0.12 | 0.04 | 0.02 | 0.05 | 0.04 | 0.09 | 0.11 | 0.07 | | 0.20 | 0.44 | 0.55 | 0.36 | 0.46 | 0.36 | 1.59 | 1.40 | 0.82 | | | | | 0.29 | 2.01 | 0.75 | 0.50 | 0.85 | 0.71 |
| 1.4 | 0.00 | 0.01 | 0.01 | | 0.01 | 0.01 | 0.01 | 0.03 | 0.08 | | 0.00 | 0.00 | 0.00 | | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | | | | | | | | | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.03 |
| 44.75 | 0.00 | 0.03 | 0.04 | | | | | | | | 0.00 | 0.00 | 0.00 | | | | | | | | 0.00 | 0.02 | 0.02 | | | | | | | | | | |
| 106.3 | 0.00 | 0.00 | 0.00 | | 0.01 | 0.00 | 0.00 | 0.08 | | | 0.00 | 0.00 | 0.00 | | 0.00 | 0.00 | 0.00 | 0.00 | | | | | | | | | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | | |
| 3.207.34 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | | | | | | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | | | | | | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | | | | | | | | |
| 5.4.79 | 0.00 | 0.00 | 0.08 | | 0.12 | 0.00 | 0.09 | 0.08 | | | 0.01 | 0.08 | 0.02 | | 0.00 | 0.00 | 0.00 | 0.00 | | | | | | | | | 0.01 | 0.03 | 0.02 | 0.03 | 0.03 | 0.02 | |
| 3.196.18 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | | 0.00 | | | | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | | | | | | 0.02 | 0.00 | 0.03 | 0.00 | 0.03 | | | 0.02 | | | | | |
| 3.64.24 | 0.00 | 0.00 | 0.00 | | 0.00 | | 0.01 | | | | 0.00 | 0.00 | 0.00 | | 0.00 | | | | | | 0.00 | 0.02 | 0.00 | | 0.03 | | | 0.10 | | | | | |
| 3.157.8 | 0.00 | 0.18 | | | | | | | | | 0.00 | 0.00 | | | | | | | | | 0.00 | 0.05 | | | | | | | | | | | |



Data Mining Discoveries

- ❖ Failed email servers
- ❖ Previously, unknown trusted relationships
- ❖ Encryption without authentication
- ❖ Possible, but unproven intrusions

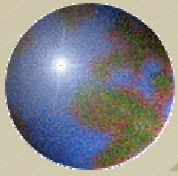


Data Mining Results

Frustrated Investigators

Frustrated Analysts

One Very Frustrated Developer



Changes to Employ Data Mining

Establish common basis of understanding

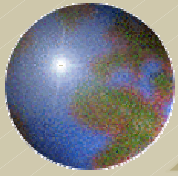
Establish criteria for reporting

- ▣ Geo-Resolution

- ▣ Timeliness

- ▣ Volume

Establish reporting procedures



Questions

Mark Kane

mkane @ ddktechgroup.com