Crunching Metrics
from Public Data Sources

Metricon 4
11 Aug 2009

Elizabeth A. Nichols Ph.D. | CTO | PlexLogic
Outline

• Context(s)
• Questions
• Metrics + Models
• Conclusions
Context: Ecosystem

Ent A
Control X

Ent B
Control Y

Ent C
Control Z
Context: Individual Business

- Breaches impact the business
- Total Impact = Impact(Expense) + Impact(MktCap)
  - Impact(expense) ~ Frequency + TotalAffected
  - Impact(mkt cap) ~ StockPrice
  - StockPrice ~ market movement + reputation

- Why should I care?
  - Stock price is clearly important. Is there any hard evidence that breaches do affect stock price, say within 60 days of breach?
  - Breach source: Inside vs Outside may imply differing controls or investments in people, process, technology
Context: Public Data Sources

Google

Olden Security Foundation

Data Loss

db
Questions

- Is Breach Frequency Increasing?
  - Frequency (Inside + Outside + Unk) is increasing
  - Frequency (Inside) is increasing
  - Frequency (Outside) is increasing

- How do Breach Frequency and Total Affected compare for Inside vs Outside sources?
  - Frequency (Inside) > Frequency (Outside)
  - TotalAffected(Inside) > TotalAffected(Outside)

- How does a breach affect stock price?
  - StockPrice(before) > StockPrice(after) for N days
  - StockPriceGradient changes sign from before(+) to after(-)
• Frequency (Inside + Outside + Unk) is increasing
• Frequency (Inside) is increasing
• Frequency (Outside) is increasing
• Frequency(Inside) > Frequency(Outside)

From the linear regression model:
Confidence > 80% that all of the above hypotheses are false.
Total Affected
Frequency Histogram

DatalossDB: log10(TotalAffected) Histogram
ds.logTA: 1Q06 to Thu Aug 6 14:23:51 2009

<table>
<thead>
<tr>
<th>parm</th>
<th>TotalAffected(TA)</th>
<th>Log10(TA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>min</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>median</td>
<td>663</td>
<td>3.48</td>
</tr>
<tr>
<td>mean</td>
<td>185,800</td>
<td>3.51</td>
</tr>
<tr>
<td>max</td>
<td>94,000,000</td>
<td>7.97</td>
</tr>
<tr>
<td>sd</td>
<td>2,614,917</td>
<td>1.23</td>
</tr>
</tbody>
</table>
Frequency by Source

Whisker Plot

* Statistically significant differences

• Outside breaches occur much more frequently than inside ones
• Inside-Accidental breaches occur much more frequently than Inside-Malicious ones.
Total Affected by Source

Whisker Plot

* Statistically significant differences

- Outside breaches affect more individuals than inside ones
- The number affected from Inside-Accidental, Inside-Malicious, and Unknown do not differ significantly
Stock Price Impact

Overlaid Sparklines: %ChgVsMkt

- 116 public companies with available, consistent ticker data
- Time Interval: [ breachDate – 30 days, breachDate + 90 days ]
- s=symbol; Price(s) = StockPrice(s);
- BreachPrice(s) = Price(s) @ BreachDate
- %Chg(s) = (Price(s) – BreachPrice(s)) / BreachPrice(s)
- %ChgVsMkt(s) = %Chg(s) - %Chg(SP500)
Interactive Symbol Picker

- Selector in upper left cell triggers updates to other three cells
- Price is sourced from [www.google.com/finance](http://www.google.com/finance)
- Proxy for SP500 is SPY
Symbol stock price is increasing in the 30 day period before and decreasing in the 90 day period after the breach.
Stock Price Impact Analysis

Gradient Distribution
Stock Price vs. Market

Statistically Significant High:
*** (>99.5%)  ** (>99%)  * (>95%)

Statistically Significant Low:
*** (>99.5%)  ** (>99%)  * (>95%)
Conclusions: Company Context

- Is Breach Frequency Increasing?
  - Frequency (Inside + Outside + Unk) is increasing \( \text{FALSE (80\%)} \)
  - Frequency (Inside) is increasing \( \text{FALSE (80\%)} \)
  - Frequency (Outside) is increasing \( \text{FALSE (80\%)} \)

- How do Breach Frequency and Total Affected compare for Inside vs Outside sources?
  - Frequency (Inside) > Frequency (Outside) \( \text{FALSE (>90\%)} \)
  - TotalAFFECTed(Inside) > TotalAFFECTed(Outside) \( \text{FALSE (>90\%)} \)

- How does a breach affect stock price?
  - StockPriceGradient changes sign from before(+) to after(-)
    - With market \( \text{TRUE (>90\%) with Market} \)
    - Against market \( \text{FALSE (>90\%)} \)
Conclusions: Ecosystem Context

- Public data can provide insight
- We are actively seeking additional data sets to analyze
- Community metrics can provide significant value via benchmarking and peer-to-peer comparisons
- Enterprise metrics programs can benefit
  - By incorporating insights from public data
  - By selective sharing of internal results
Appendix: Statistical Analysis

- 16 possible observations; 116 companies
- Hypotheses:
  - H0: Pr(observaation) <= 1/16
  - H1: Pr(observation) > 1/16
- n=# obs=116; p0=1/16, q0=(1-p)=15/16; npq = 6.9
- Npq>5 → Use normal distribution as an approximation to the binomial distribution for hypothesis testing
- \[ Z = \frac{p - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \] yeilds the # sd’s from mean for each obs.
Appendix: p-values

<table>
<thead>
<tr>
<th>Gradients</th>
<th># Obs</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>--:--</td>
<td>11</td>
<td>.08</td>
</tr>
<tr>
<td>--:+</td>
<td>3</td>
<td>.95</td>
</tr>
<tr>
<td>--:+</td>
<td>6</td>
<td>.69</td>
</tr>
<tr>
<td>--:++</td>
<td>4</td>
<td>.54</td>
</tr>
<tr>
<td>-+:--</td>
<td>7</td>
<td>.15</td>
</tr>
<tr>
<td>-+:+</td>
<td>10</td>
<td>.08</td>
</tr>
<tr>
<td>-+:++</td>
<td>1</td>
<td>.99</td>
</tr>
<tr>
<td>-+:++</td>
<td>11</td>
<td>.69</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gradients</th>
<th># Obs</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>+:--</td>
<td>6</td>
<td>.69</td>
</tr>
<tr>
<td>++:+</td>
<td>2</td>
<td>.98</td>
</tr>
<tr>
<td>+:++</td>
<td>13</td>
<td>.01</td>
</tr>
<tr>
<td>+:+:</td>
<td>11</td>
<td>.08</td>
</tr>
<tr>
<td>++:--</td>
<td>1</td>
<td>.99</td>
</tr>
<tr>
<td>++:+</td>
<td>6</td>
<td>.69</td>
</tr>
<tr>
<td>++:+:</td>
<td>9</td>
<td>.26</td>
</tr>
<tr>
<td>++:++</td>
<td>16</td>
<td>.0005</td>
</tr>
</tbody>
</table>

The p-value indicates the probability of rejecting a true H0, namely that the probability of any of the above possible observations is 1/16 or less.
Elizabeth A. Nichols, Ph.D.
CTO for Metrics, PlexLogic
betsy.nichols@plexlogic.com
+1 (703) 963-7202

http://www.metricscenter.org