

**DEFCON XI Session**  
**"Online Corporate Intelligence"**

NOTES TO CONFERENCE GOER/READER:

I'm never quite sure what to provide when the DEFCON organizers ask for materials to be included in the conference CD ROM. Especially where there are no obvious materials to include (like scripts). Initially, I was going to include a PowerPoint presentation for distribution, but I decided not to because PowerPoint presentations deserve narration and are ineffective without it.

So instead, I'm providing a RTF formatted file with the rough text of what I anticipate my final slides to contain, with a brief narration to the right.

The final DEFCON XI PowerPoint slides will have the same general information contained here, but with better presentation. I encourage you to view the final presentation video on the DEFCON web site, if--or when--it becomes available, as others have in the past.

Thank you, and see you in Vegas!  
Michael Schrenk

July 7, 2003

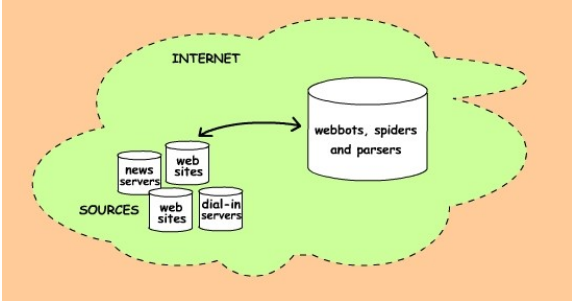
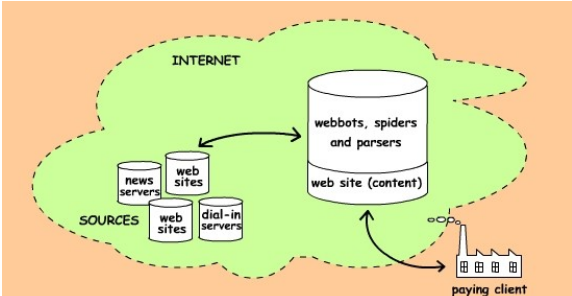
mike@schrenk.com  
www.schrenk.com

Planned Slide (content)	Brief Narration
<p data-bbox="256 321 724 348"><b>Online Corporate Intelligence</b></p> <p data-bbox="289 384 691 499"><i>Knowing Your Competition, Protecting Your Name &amp; Understanding Your Markets</i></p>	<p data-bbox="833 254 1320 348">A rapidly growing number of businesses use webbots and spiders to collect corporate intelligence about their competitors and competitive markets.</p>
<p data-bbox="172 653 464 680"><b>What will we cover</b></p> <ul data-bbox="220 682 768 892" style="list-style-type: none"> <li data-bbox="220 682 768 741">• How online intelligence differs from traditional methods</li> <li data-bbox="220 743 768 802">• The difference between intelligence and espionage</li> <li data-bbox="220 804 768 831">• Corporate "dash boards"</li> <li data-bbox="220 833 768 861">• Tips from the field</li> <li data-bbox="220 863 768 892">• Opportunities for the community</li> </ul>	<p data-bbox="833 646 1320 741">Online intelligence gathering has some distinct differences from traditional methods of gathering intelligence about business competition.</p> <p data-bbox="833 766 1320 861">The focus of this session is on using automated online techniques to collect information about business competitors and competitive markets.</p> <p data-bbox="833 886 1320 1075">Providing online intelligence services to organizations is a new area for developers who also possess modest hacking skills. And, when done within the law, is completely legal and distinguishes a select group of developers from the hordes of other developers.</p>

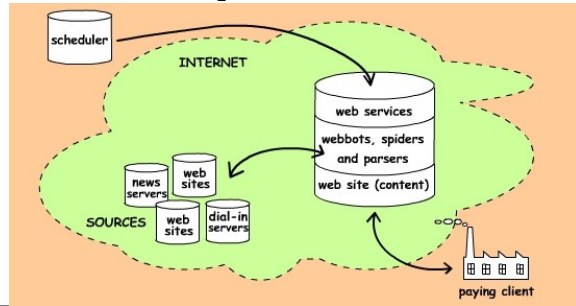
<p><b>Who am I?</b></p> <ul style="list-style-type: none"> <li>• DEFCON X: "Introduction to Writing Spiders and Agents"</li> <li>• Write webbots and spiders for corporate clients</li> <li>• Minneapolis-based consultancy of Michael Schrenk Ltd.</li> <li>• Also write, speak and teach.</li> </ul> <p><b>mike@schrenk.com    www.schrenk.com</b></p>	<p>I have held various executive and consulting positions in the online industry and now head the consultancy of Michael Schrenk Ltd., whose primary mission is to write webbots that create competitive advantages for companies by collecting and providing context for real time intelligence from the Internet and other digital sources.</p> <p>A past DEF CON speaker; I have also created <b>traditional</b> online applications for Disney, Nike, Adidas and Callaway Golf and has written for Computerworld and Web Techniques magazines.</p> <p>I hold a patent for mobile server technology and have a process patent pending for online auctions.</p> <p>More information is available at <a href="http://www.schrenk.com">www.schrenk.com</a></p>
<p><b>Intelligence = information</b></p> <ul style="list-style-type: none"> <li>• What can you find out about your market (trends)?</li> <li>• What can you find out about your competition?</li> <li>• What do people know about you?</li> </ul>	<p>(Self-explanatory, no narration needed.)</p>

<p style="text-align: center;"><b>Definitions</b></p> <p>Intelligence is not necessarily:</p> <ul style="list-style-type: none"> <li>• Espionage, or</li> <li>• A covert action</li> </ul>	<p>These are my definitions and they may not fit the criteria you'd find in Webster, but I think it's important for the sake of framing the discussion. Defining what you're doing--and having guidelines-- can also keep one out of trouble.</p> <p><u>Espionage</u>: Spying, bugging, illegally obtaining information. There's so much information online there's no need to resort to espionage.</p> <p><u>Covert action</u>: A covert action is intended to influence the outcome of a strategy or vote.</p> <p>You "might" say that sniping software fits the description of a covert action. (More on sniping later.)</p>
<p><b>Gathering intelligence means learning new habits</b></p> <ul style="list-style-type: none"> <li>• To get the most corporate intelligence from the Internet you should use: <ul style="list-style-type: none"> <li>- A browser?</li> <li>- An email client?</li> <li>- A newsreader?</li> <li>- Telnet? Or</li> <li>- Other...</li> </ul> </li> </ul>	<p>We are conditioned to think along certain lines of thought. For many people--because the Web is the only protocol they use, see the World Wide Web as synonymous with the Internet, which it isn't.</p> <p>The web agent we use defines the way we use the Internet.</p> <p>Doing what everyone else does provides no competitive advantage.</p> <p>There's no reason to perform online corporate intelligence if it doesn't give your client a competitive advantage.</p> <p>To achieve a competitive advantage you'll need to do things that your competition isn't doing.</p>
<p><b>Traditional methods of (legally) collecting corporate intelligence</b></p> <ul style="list-style-type: none"> <li>• Conferences &amp; sales literature</li> <li>• Employee trading</li> <li>• Patent records</li> <li>• Secret shoppers</li> <li>• Help Wanted Ads</li> </ul>	<p>People don't think a lot about employee trading, but it happens all the time when people change jobs in cities that have more than one major player in a particular industry.</p> <p>Portland: Nike, Adidas  Detroit: GM, Ford, Chrysler  Orlando: Universal Studios, Disney, Sea Land</p> <p><i>This also happens, of course, when major companies are not in the same city, but with less frequency.</i></p>

<p><b>Disadvantages to traditional corporate intelligence collecting</b></p> <ul style="list-style-type: none"> <li>• Requires contact</li> <li>• Mostly one-time activities</li> <li>• Cannot be done anonymously</li> <li>• Expensive</li> </ul>	<p>Self-explanatory, no narration needed.</p>
<p><b>What distinguishes online corporate intelligence?</b></p> <ul style="list-style-type: none"> <li>• Can be done from a distance (with stealth)</li> <li>• Can be automated</li> <li>• Can be done anonymously (for the most part)</li> <li>• Can be interactive and/or proactive</li> <li>• Reduces or eliminates latency between when an event happens and when a decision can be made.</li> <li>• Can <b>create</b> relevance that traditional methods can't</li> </ul>	<p>You will notice that many of these characteristics are the same as the reasons why computer hacking is so dangerous to organizations.</p> <p>Possibility the most interesting factor here is that the same mechanisms for collecting online intelligence can also:</p> <ol style="list-style-type: none"> <li>1. Providing context or relevance (more of this in dashboard section), and</li> <li>2. Be proactive or interactive, and essentially "think" for us.</li> </ol>
<p>Creating relevance</p> <ul style="list-style-type: none"> <li>• Online <ul style="list-style-type: none"> <li>○ Cross reference multiple sources</li> <li>○ Can be updated over time</li> </ul> </li> <li>• <b>Time</b> can provide context to public information <ul style="list-style-type: none"> <li>○ When associated with time you can show trends</li> </ul> </li> <li>• <b>Relationships</b> between data <ul style="list-style-type: none"> <li>○ Can provide added context for data</li> </ul> </li> </ul> <p><i>Online applications can automate the evaluation process.</i></p>	
<p><b>Sources for intelligence gathering (part 1)</b></p> <ul style="list-style-type: none"> <li>• Corporate web sites <ul style="list-style-type: none"> <li>○ Job postings</li> <li>○ Product pricing</li> <li>○ News</li> </ul> </li> </ul>	<p>(Self-explanatory, no narration needed.)</p>
<p><b>Sources for intelligence gathering (part 2)</b></p> <ul style="list-style-type: none"> <li>• Government web sites <ul style="list-style-type: none"> <li>○ Court records</li> <li>○ SEC filings</li> <li>○ Patent records</li> <li>○ Census data</li> </ul> </li> </ul>	<p>(Self-explanatory, no narration needed.)</p>

<p><b>Sources for intelligence gathering (part 3)</b></p> <ul style="list-style-type: none"> <li>• Online auctions <ul style="list-style-type: none"> <li>◦ Interactive webbots</li> </ul> </li> <li>• Whois servers</li> <li>• News servers</li> <li>• HTTP headers</li> </ul>	<p>(Self-explanatory, no narration needed.)</p>
<p><b>Technology</b></p> <ol style="list-style-type: none"> <li>1. Gather, parse and store data</li> <li>2. Organize (provide context) and make available on the web</li> <li>3. Provide a trigger for the webbots</li> </ol>	<p>The next three slides show how I typically like to set up the technology for delivering online corporate intelligence.</p>
<p><b>Identify sources, write bots, store data</b></p>  <p>The diagram shows a green cloud labeled 'INTERNET'. Inside the cloud, on the left, are four cylinders representing 'SOURCES': 'news servers', 'web sites', 'web sites', and 'dial-in servers'. An arrow points from these sources to a larger cylinder on the right labeled 'webbots, spiders and parsers'.</p>	<p>Once sources of information have been identified the next step is to write or obtain the software required to download and parse data and store data in a database.</p> <p>Your sources don't have to be web sites, they can include any digital medium that's available via a network including dial-up services, ftp and news servers. In fact, your data--when combined--is apt to be stronger if it comes from a variety of protocols.</p> <p>If you're not doing any dial-up connections to main-frames, etc., the server used for these systems can be a simple ~\$10/month web site that supports the technology you use, which in my case is PHP/cURL/MySQL.</p> <p>If you are doing dial-up connections you'll need a local server, or at least one with access to a phone line.</p>
<p><b>Write a data driven web site for the customer of the data</b></p>  <p>The diagram is similar to the previous one, but the cylinder on the right is now labeled 'webbots, spiders and parsers' and 'web site (content)'. An arrow points from this cylinder to a small building icon labeled 'paying client'.</p>	<p>Once the sources are identified and the data parsed and stored in a database, a web site (with authentication) can be created to allow a paying customer the ability to view the data.</p> <p>I like supplying data in this manner because it gives the developer more control than just emailing the data to the customer.</p>

**Create a scheduled trigger for the webbots and spiders**



The final step is to provide some type of trigger to tell the webbots, etc. when to run.

An easy way to do this is with a web service that tells the webbots when to run and what to collect.

A simple desktop computer periodically running a scheduled service or a CHRON job can invoke the web service.

**Corporate intelligence "dashboards"**

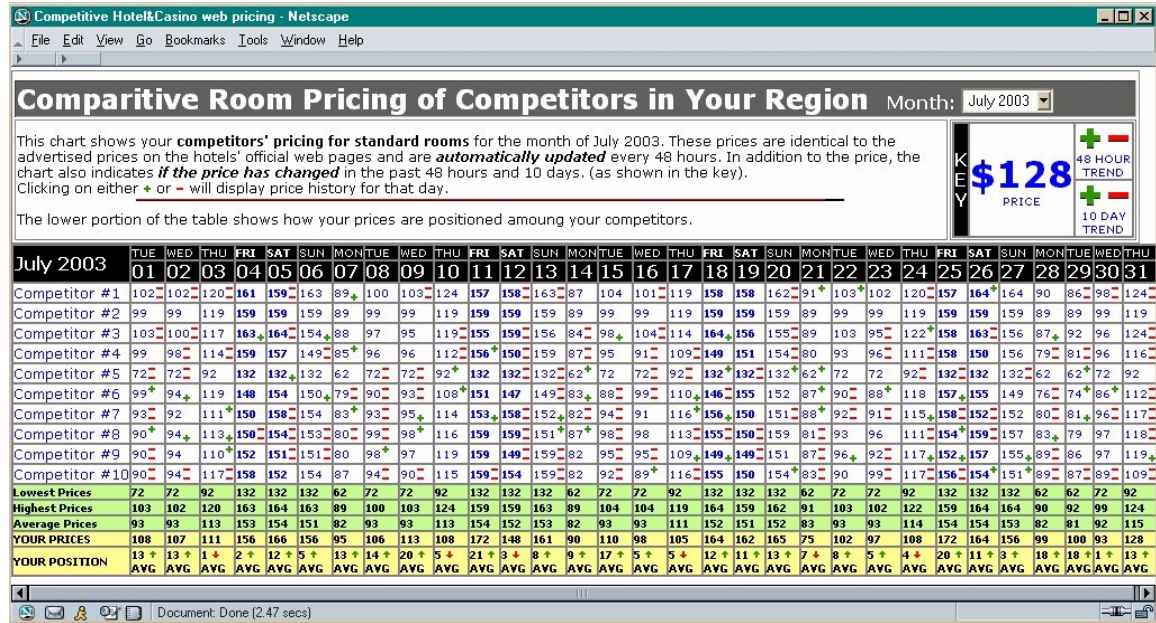
- Provide "big picture" of data within context
- Adds context to data
  - Context filters
  - Context via math
  - Context via time
- Creates branding opportunities

The idea of a dashboard is to:

1. Create context for data
2. Combine data from many sources
3. Make data available in real time

This particular dashboard shows comparative hotel room prices for a client and ten of its competitors. This intelligence information is helpful for:

1. Determining a truly competitive price
2. By using the accompanying statistics, it definitively shows how the clients room prices are aligned with competing hotel prices.
3. Predicting market pricing trends



You can add even more relevance to this data by combining it with other sources. What kinds of value could one add to this data if it also included statistics that include:

1. Occupancy rates (both your client's and client's competitors')
2. Profits and losses
3. Changes relative to last year/quarter/month
4. Extraneous influences (enlarged casino, road construction, employee changes, process updates, etc.)

### Another Example:

#### Policing the Internet

##### Problem:

1. People steal things and want to liquidate quickly.
2. Online auctions are an attractive alternative to pawn shops

##### Solution:

1. Create a online interface that allows law enforcement to enter groups of items that were stolen at the same time
2. Write webbots that look for individuals who are selling the same grouping of items on eBay.

Part of me is sad that I decided to title this session "CORPORATE" Intelligence because many non-corporate entities can also benefit from online intelligence (as seen to the left).



**Example:**  
**Study what people are studying**

- **Amazon Purchase Circles**



Back when my job was to evaluate emerging technologies for a MN medical technology company, I would use the "purchase circles" that are available on Amazon.com to track what new books our competitors were reading.

Amazon's purchase circles are collections of "most ordered" books that emanate from particular IP address blocks. A simple reverse IP look-up can reveal where those orders are coming from. Amazon publishes this information as a means for readers to see what's popular in particular localities or within particular industries.

With reading circles you can tell what's popular with various corporations, government branches, or educational institutions.

I'm surprised to see them again, because they were removed (I think) at one point because people saw it as an invasion of privacy.

**Amazon Reading Circles**

- Top books Apple Computer ordered from Amazon
  - 1 Mac OS X in a Nutshell
  - 2 Mac OS X Hacks
  - 3 Mac OS X: The Missing Manual, Second Edition
  - 4 Pattern Recognition
  - 5 Harry Potter and the Order of the Phoenix
  - 6 What Should I Do with My Life?

You would expect a technical company like Apple to have some technical books in its list of most popular Amazon purchases.

What I would look for here are anomalies, like the book on pattern recognition. Why would this book be this high up on the list? Is there other information that could be gathered to reinforce the validity or shine more light on this finding?

These lists can definitely help define corporate culture

**Example:**  
**Find out who companies are hiring**

- Watching corporate help wanted ads can expose strategies
- Collecting this data with a webbots and applying statistical methods can reviled trends

(Self-explanatory, no narration needed.)

<p><b>Example:</b>  <b>Interactive intelligence Part 1.</b></p> <p><b>Sniping software:</b></p> <ul style="list-style-type: none"> <li>• Software that places last second bids on online auctions.</li> <li>• Prevents the bidding process from raising auction prices</li> <li>• Somewhat limited by proxy bidding.</li> </ul>	<p>(Self-explanatory, no narration needed.)</p>
<p><b>Example:</b>  <b>Interactive intelligence Part 2.</b></p> <p><b>Intelligent shopping software:</b></p> <ul style="list-style-type: none"> <li>• Webbots collect market information on select items</li> <li>• Purchases may be made automatically when specific criteria (based on collected data) are met</li> <li>• Stocks, online auctions, etc.</li> </ul>	<p>(Self-explanatory, no narration needed.)</p>
<p><b>Example:</b>  <b>Online clipping services</b></p> <ul style="list-style-type: none"> <li>• Webbots and spiders look for information about competitors.</li> <li>• Can also be used to see what people are saying about you.</li> </ul>	
<p><b>Why you need to treat online sources respectfully</b></p> <ul style="list-style-type: none"> <li>• They can turn you away</li> <li>• They can discover what you're doing</li> </ul>	<p>(Self-explanatory, no narration needed.)</p>

<p><b>How to treat sources kindly</b></p> <ul style="list-style-type: none"> <li>• Respect Bandwidth</li> <li>• Introduce Randomness</li> <li>• Respect Privacy</li> </ul>	<p><b>Bandwidth:</b>  Every time you hit a server you leave a record in a server log file. If you hit that server every five seconds you will not only become very prominent in the log files, but also very suspicious looking.</p> <p><b>Randomness:</b>  Randomness in the way your webbots schedule their tasks and use login credentials can make your innocent dealings less suspect looking. Even if you are not doing anything nefarious, if you look sneaky, they may justify blocking your traffic.</p> <p><b>Privacy:</b>  Just because something is publicly available doesn't mean it needs to be communicated recklessly.</p> <p>There's a lot of public information that probably <i>shouldn't be public</i>. For example, there are still many places online to get social security numbers.</p> <p>I highly recommend that you treat information with respect and don't do anything rash. The more disrespectful you are of information, the greater the odds of getting into trouble.</p>
<p><b>Tips on writing stealthy webbots and agents</b></p> <ul style="list-style-type: none"> <li>• Bandwidth considerations</li> <li>• Randomness <ul style="list-style-type: none"> <li>◦ Time</li> <li>◦ Order</li> </ul> </li> <li>• Rotating IP addresses</li> <li>• Using a link proxy <ul style="list-style-type: none"> <li>◦ Required in a "dash board" with outside links</li> <li>◦ Destroys the referrer HTTPd variable</li> </ul> </li> </ul>	<p><b>Link Proxies</b>  If you have an outside link in a dashboard, or some other intelligence collection, you risk the chance that the HTTP_REFERER server variable will reveal the existence of your intelligence cache.</p> <p>A link proxy basically bounces the hyper reference off of a server (or two and strips the REFERER variable off the request header.</p>
<p style="text-align: center;"><b>Thank you</b></p>	<p style="text-align: center;">mike@schrenk.com  www.schrenk.com</p>