Are You Raising Your Internet Assets in a Bad Neighborhood?

Andrew Hay, CISO
DataGravity, Inc.
ahay@datagravity.com
@andrewsmhay
<table>
<thead>
<tr>
<th></th>
<th>Agenda</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>About</td>
</tr>
<tr>
<td>2</td>
<td>Overview &amp; Methodology</td>
</tr>
<tr>
<td>3</td>
<td>Data Collection</td>
</tr>
<tr>
<td>4</td>
<td>Costs</td>
</tr>
<tr>
<td>5</td>
<td>Conclusion / Q&amp;A</td>
</tr>
</tbody>
</table>
About Andrew Hay

• Andrew Hay
  – Chief Information Security Officer (CISO) @ DataGravity

• Former:
  – Director of Research @ OpenDNS
  – Chief Evangelist & Director of Research @ CloudPassage
  – Senior Security Analyst @ 451 Research
  – Sr. Security Analyst in higher education and a bank in Bermuda

• Wrote some books, blog, spend more time on planes than I care to mention...
Agenda

1. About
2. Overview & Methodology
3. Data Collection
4. Costs
5. Conclusion / Q&A
Everyone wants to believe that their cloud workloads are safe, secure, and isolated.
The reality is, however, that clouds are multitenant and you have no idea of, or control over, who your neighbors are.
If You Were Looking to Move Into A New Apartment Complex
But You Found Out That It Was Full of Crack Heads?
Really Full...
But You Moved In Anyway…
But You Moved In Anyway…
What Would Your Mother Think?
What Would Your Mother Think?
What Would Your Mother Think?

My son’s a crack head!!!!
What If This Was a Really Bad Neighborhood?

What is my son thinking?!?!!
Or a Really Bad Part of Town?

Well, at least his sister isn’t an idiot...
Bad Neighborhood Model
Bad Neighborhood Model
Bad Neighborhood Model
Bad Neighborhood Model

- Region
  - Neighborhood
    - Apartment Building
      - Unit
We Have Ways of Checking These Things

We can ask friends, family, strangers

We can do our own research and review crime statistics

- **FBI** - [http://www.fbi.gov/stats-services/crimestats](http://www.fbi.gov/stats-services/crimestats)
  - Uniform Crime Reports, Hate Crime Statistics, Law Enforcement Officers Killed and Assaulted (LEOKA), National Incident-Based Reporting System (NIBRS)

- **Bureau of Justice Statistics (BJS)** - [http://www.bjs.gov](http://www.bjs.gov)
  - National Crime Statistics Exchange (NCS-X)

- **Neighborhood Scout** - [http://www.neighborhoodscout.com](http://www.neighborhoodscout.com)

- **CrimeReports** - [https://www.crimereports.com](https://www.crimereports.com)
Manhattan vs. Newark
San Jose vs. San Francisco
Why Don’t We Think This Way About Hosting

You would typically perform due diligence on the neighborhood before:

- Moving your family to a new city or suburb…
- Taking a new job in a city you’ve never lived in before…
- Moving your business to, or expanding into, a new neighborhood…

Shouldn’t we perform similar due diligence of the Internet ‘neighborhood’ we’re looking to ‘move in’ to?
Bad Internet Neighborhood Model
Bad Internet Neighborhood Model
Bad Internet Neighborhood Model

Prefix / Subnet

ASN

IP Address
Consequences of A Bad Neighborhood

- Crime rate
- Property value
- Desire to stay
- Desire to leave
- Vacancy rate
Consequences of A Bad Internet Neighborhood

- Risk of having IP/Prefix blocked by customers
- Network/system latency (DDoS/scanning)
- Risk of collateral damage (hack back)
- Risk of service disruption by search & seizure
How Are We Choosing Hosting Providers Today?

“Up, and to the right”
How Are We Choosing Hosting Providers Today?

“Up, and to the right”
How Are We Choosing Hosting Providers Today?

*Lowest price available*
How Are We Choosing Hosting Providers Today?

Lowest price available
Race to the bottom on hourly/monthly rates

How Are We Choosing Hosting Providers Today?

*It’s certified, so it must be good/secure/safe*
How Are We Choosing Hosting Providers Today?

Feature/functionality

<table>
<thead>
<tr>
<th>AWS PRODUCT CATEGORIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview</td>
</tr>
<tr>
<td>Compute</td>
</tr>
<tr>
<td>Storage &amp; Content Delivery</td>
</tr>
<tr>
<td>Databases</td>
</tr>
<tr>
<td>Networking</td>
</tr>
<tr>
<td>Administration &amp; Security</td>
</tr>
<tr>
<td>Analytics</td>
</tr>
<tr>
<td>Application Services</td>
</tr>
<tr>
<td>Deployment &amp; Management</td>
</tr>
<tr>
<td>Mobile Services</td>
</tr>
<tr>
<td>Enterprise Applications</td>
</tr>
<tr>
<td>AWS Support</td>
</tr>
<tr>
<td>AWS Marketplace Software</td>
</tr>
</tbody>
</table>
How Are We Choosing Hosting Providers Today?

**Feature/functionality**

<table>
<thead>
<tr>
<th>Compute</th>
<th>Storage</th>
<th>Networking</th>
<th>Big Data</th>
<th>Services</th>
<th>Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compute Engine</td>
<td>Storage</td>
<td>Load Balancing</td>
<td>BigQuery</td>
<td>Translate API</td>
<td>Deployment Manager</td>
</tr>
<tr>
<td>App Engine</td>
<td>Bigtable</td>
<td>Interconnect</td>
<td>Dataflow</td>
<td>Prediction API</td>
<td>Logging</td>
</tr>
<tr>
<td>Container Engine</td>
<td>Datastore</td>
<td>DNS</td>
<td>Pub/Sub</td>
<td>Endpoints</td>
<td>Monitoring</td>
</tr>
</tbody>
</table>

**AWS PRODUCT CATEGORIES**

- Overview
- Compute
- Storage & Content Delivery
- Databases
- Networking
- Mobile Services
- Enterprise Applications
- AWS Support
- AWS Marketplace Software
How Are We Choosing Hosting Providers Today?

Feature/functionality

<table>
<thead>
<tr>
<th>Compute</th>
<th>Storage</th>
<th>Networking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compute Engine</td>
<td>Storage</td>
<td>Load Balancer</td>
</tr>
<tr>
<td>App Engine</td>
<td>Bigtable</td>
<td>Interconnect</td>
</tr>
<tr>
<td>Container Engine</td>
<td>Datastore</td>
<td>DNS</td>
</tr>
</tbody>
</table>

Overview

AWS PRODUCT CATEGORIES

- Overview
- Compute
- Storage & Content Delivery
- Databases
- Networking

Services

- Translate API
- Prediction API
- Endpoints

Management

- Deployment Manager
- Logging
- Monitoring

Mobile Services
- Enterprise Applications
- AWS Support
- AWS Marketplace Software
Can We Do Better?
How Did We Get The Data?

• Lots of cloud providers out there but we decided to start with AWS due to its popularity
How Did We Get The Data?

- Lots of cloud providers out there but we decided to start with AWS due to its popularity
- Identified the ASNs

http://bgp.he.net/search?search%5Bsearch%5D=amazon&commit=Search
How Did We Get The Data?

- Lots of cloud providers out there but we decided to start with AWS due to its popularity
- Identified the ASNs
- Counted number of Prefixes (a.k.a. CIDRs)

1 Providers
14 ASNs
2,110 Prefixes

How Did We Get The Data?

- Lots of cloud providers out there but we decided to start with AWS due to its popularity
- Identified the ASNs
- Counted number of Prefixes (a.k.a. CIDRs)
- Identified distinct number of hosting countries

http://dev.maxmind.com/geoip/geoip2/geolite2/
How Did We Get The Data?

- Lots of cloud providers out there but we decided to start with AWS due to its popularity
- Identified the ASNs
- Counted number of Prefixes (a.k.a. CIDRs)
- Identified distinct number of hosting countries
- Determined count of total useable IP addresses per Prefix

1 Providers
14 ASNs
2,110 Prefixes
9 Countries
10,130,200 IP Addresses

How Did We Get The Data?

- Lots of cloud providers out there but we decided to start with AWS due to its popularity
- Identified the ASNs
- Counted number of Prefixes (a.k.a. CIDRs)
- Identified distinct number of hosting countries
- Determined count of total useable IP addresses per Prefix
- Count the number of malicious domains identified and blocked by various services

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Providers</td>
<td>1</td>
</tr>
<tr>
<td>ASNs</td>
<td>14</td>
</tr>
<tr>
<td>Prefixes</td>
<td>2,110</td>
</tr>
<tr>
<td>Countries</td>
<td>9</td>
</tr>
<tr>
<td>IP Addresses</td>
<td>10,130,200</td>
</tr>
<tr>
<td>Malicious Domains</td>
<td>???</td>
</tr>
</tbody>
</table>
Data Enrichment
Agenda

1. About
2. Overview & Methodology
3. Data Collection
4. Costs
5. Conclusion / Q&A
Cost Data

AWS provides a Price List API

http://docs.aws.amazon.com/awsaccountbilling/latest/aboutv2/price-changes.html

Need to consider:

• Hourly vs. reserved instance cost
• Region specific pricing
  • e.g. US-East is less expensive than US-West

$TBD Cost/Hour

$TBD Cost/Month

$TBD Cost/Year
Required Algorithms

**Similar to Zillow Zestimate®**


- **Physical attributes:** Location, lot size, square footage, number of bedrooms and bathrooms and many other details.
- **Tax assessments:** Property tax information, actual property taxes paid, exceptions to tax assessments and other information provided in the tax assessors' records.
- **Prior and current transactions:** Actual sale prices over time of the home itself and comparable recent sales of nearby homes.

“We use proprietary automated valuation models that apply advanced algorithms to analyze our data to identify relationships within a specific geographic area, between this home-related data and actual sales prices. Home characteristics, such as square footage, location or the number of bathrooms, are given different weights according to their influence on home sale prices in each specific geography over a specific period of time, resulting in a set of valuation rules, or models that are applied to generate each home’s Zestimate.”
Agenda

1. About
2. Overview & Methodology
3. Data Collection
4. Costs
5. Conclusion / Q&A
Features For Models

- IP
- CIDR
- ASN
- Region
- Cloud

Values
Counts
Stats
Scores
The presence of neglected applications and improperly configured servers should have an adverse effect on hosting provider valuation.

Careful consideration when selecting a particular hosting provider, in addition to the country, region, and even the subnet within which your guest instance resides, is now warranted.
Conclusions

Statistical models for determining hosting provider valuation in concert with discovered vulnerabilities, third-party intelligence enrichment, and infrastructure availability **MUST** be developed
Questions…or Ideas?

Are You Raising Your Internet Assets in a Bad Neighborhood?

Andrew Hay, CISO
DataGravity, Inc.
ahay@datagravity.com
@andrewsmhay