Machine-Learning and Big-Data Research on Anomaly Identification

Neil C. Rowe

U.S. Naval Postgraduate School
Monterey, California, USA
ncrowe@nps.edu
http://faculty.nps.edu/ncrowe
We have too much data

• Increasing storage capacities are permitting storage of more and more data.

• Increasing networking means more and more access to data – but coordination is required.

• Unfortunately, many traditional methods of data processing do not scale up: Handling 1000 times more data often takes more than 1000 times as much time.

• Errors are also increasing as much data is automatically generated and is not checked by people.

• Big data is not necessarily centralized.
Big data offers new opportunities

• Automated big data may see trends faster. Example: Google can predict flu epidemics from query terms before doctors know about them.

• Traditionally, analysis took samples of data and analyzed them. But if we can look at all the data, we may see rare trends missing in a sample.

• Example: Analyzing indicators of an adversary’s military or political intentions. They can try to conceal them, but something may slip through in a large amount of data.
Increasing processing capabilities make machine learning possible with big data

• The speed of computers has increased steadily over the last forty years. Problems that seemed too hard to do by computer a few years ago are now doable.

• Storage capacities have greatly increased too. “Big data” learning has led to impressive advances recently.

• Machine learning requires much time and storage because it requires some guessing and blind alleys.

• Good instruction can reduce this, but learning is hard for everyone. People learn many wrong things from experience (hence prejudice).
Projects discussed in this talk

• Identifying anomalies in aircraft and ship tracks
• Finding new cyberattacks on digital systems using honeypots
• Tracking people who are learning physical-motion tasks
Identifying anomalies in aircraft tracks
Big data for aircraft track analysis

• Analysis of what aircraft are doing can be hard:
  • Aircraft move quickly.
  • They may not aid identification.
  • They may be subject to system malfunctions, or may be misconfigured, so they may not self-identify.
  • They are less often in airlanes, especially if they are autonomous.

• Nonetheless, we now have basic aircraft tracking data from satellite coverage.

• We can use this to analyze aircraft as a big-data problem.

• Since militaries have limited bandwidths, our strategy is to push some of the processing and intelligence “to the edge” or to the platforms that collect data.
We have two challenges: Report key data to a central command node, and report a different set of key data to sibling aircraft. There’s not enough bandwidth to report all data to either place. We need to focus on “interesting” aircraft.
Experimental setup

• Our experiments used unclassified air traffic-control data from the ADS-B database for testing.
• This data is obtained from satellites.
• It averaged about 1.6 gigabytes and 11.3 million records of data per day for most aircraft in the air around the world.
• For testing we extracted data from the 10th day of each month from May 2015 through April 2016 so as reduce seasonal effects.
• The attributes selected for anomaly analysis were ICAO code of the aircraft if known, type of the aircraft, operator, country of origin, altitude, latitude, longitude, speed, bearing, and timestamp.
• Some of the data was clearly faulty and were deleted.
Aircraft heading and timestamp patterns
Aircraft speed versus altitude
Our data flow for anomaly ranking

- Inferred tracks and their averages
- Surveillance priorities: spacetime, type, event
- Command data aggregation
- Aircraft data, currently from ADS-B
- Ranked list of most anomalous data
- Spacetime-filtered data for sibling platforms
- Lat/long averages
- Speed-altitude averages
- Hour of day histogram
- Aircraft-owner histogram
- Lat/long/day histograms
## Variations in owner and type by date

<table>
<thead>
<tr>
<th>Aircraft registration and type</th>
<th>Total of 10(^{th}) day of all 12 months</th>
<th>September 10, 2015</th>
<th>March 10, 2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>North America commercial</td>
<td>31</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>South and Central America commercial</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Europe commercial</td>
<td>3851</td>
<td>615</td>
<td>201</td>
</tr>
<tr>
<td>Africa commercial</td>
<td>144</td>
<td>28</td>
<td>21</td>
</tr>
<tr>
<td>Middle East commercial</td>
<td>588</td>
<td>62</td>
<td>212</td>
</tr>
<tr>
<td>Central Asia commercial</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>East Asia and Pacific commercial</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>North America noncommercial</td>
<td>16</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Other America noncommercial</td>
<td>75</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Europe noncommercial</td>
<td>678</td>
<td>249</td>
<td>2</td>
</tr>
<tr>
<td>Africa noncommercial</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Middle East noncommercial</td>
<td>13</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Central Asia noncommercial</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>East Asia and Pacific noncommercial</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Track analysis

Left plot is aircraft positions southeast of San Francisco airport; right plot is track endpoints in the same area. Anomalies are more clear on the right.
Aircraft anomaly measures we tested

- Count in its latitude/longitude bin
- Typicality of the commercial/noncommercial identification in the bin
- Typicality of its heading in the bin
- Typicality of its altitude in the bin
- Typicality of its speed in the bin
- Typicality of the time of day
- Traffic rate of the particular day
- Degree to which the aircraft is in an airlane
- The deviation of the lat/long path from a straight path
- The average turn rate of the aircraft
- The average altitude change
- The atypicality of the count on this day in this lat/long bin
- The atypicality of the type distribution in this bin on this day
- The atypicality of the speed distribution in this bin on this day
Histogram of the values of anomaly factors
Principal components of a random data sample

This suggests another way of finding anomalies: Cluster, look for points outside the clusters.
Altitude deviation versus great-circle deviation

Outliers are interesting aircraft.
Histogram of ship-anomaly measures

Anomaly factor histogram

- count_ratio_average_difference
- direction_average_difference
- spd_average_difference
- log_op_freq_average_difference
- hour_average_anomaly_difference
- log_count_rarity_difference
- path_variance_difference
- day_count_deviation_difference
- difference_sum

Anomaly strength
## HDFS processing times

<table>
<thead>
<tr>
<th>Process</th>
<th>Time for single processor on the Spark site</th>
<th>Time for Hadoop Spark implementation (28 processors)</th>
<th>Time per record for Spark implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Move data from database to files</td>
<td>612 minutes</td>
<td>26.6 minutes</td>
<td>0.00117353</td>
</tr>
<tr>
<td>Count aircraft in lat-long bins</td>
<td>70.6 minutes</td>
<td>2.5 minutes</td>
<td>0.00011029</td>
</tr>
<tr>
<td>Average the speed in altitude bins</td>
<td>72.7 minutes</td>
<td>1.0 minutes</td>
<td>0.00004412</td>
</tr>
<tr>
<td>Summarize the aircraft tracks</td>
<td>59.5 minutes</td>
<td>22.3 minutes</td>
<td>0.00098382</td>
</tr>
<tr>
<td>Find anomalies using the first 7 factors</td>
<td>211.2 minutes</td>
<td>18.5 minutes</td>
<td>0.00081617</td>
</tr>
</tbody>
</table>
## Processing times on a 2000-core supercomputer

<table>
<thead>
<tr>
<th>Aircraft-data Process</th>
<th>Total time for Hadoop/Gradle on supercomputer, 1.5 billion records, 312 processors</th>
<th>Time per record</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set up processing</td>
<td>33.25 minutes</td>
<td>0.00000130</td>
</tr>
<tr>
<td>Aggregate the data</td>
<td>9.60 minutes</td>
<td>0.00000038</td>
</tr>
<tr>
<td>Find anomalies using the first 10 aircraft factors</td>
<td>10.27 minutes</td>
<td>0.00000040</td>
</tr>
<tr>
<td>Assign clusters based on cluster centers found previously in uniprocessor implementation</td>
<td>112.87 minutes</td>
<td>0.000000442</td>
</tr>
<tr>
<td>Create graphs of anomaly analysis (optional)</td>
<td>53.18 minutes</td>
<td>0.00000208</td>
</tr>
</tbody>
</table>
Conclusions about aircraft anomalies

• Most routine aircraft data are very predictable, so anomalies stand out well.

• A linear model of anomaly factors worked well. This suggests a nonlinear model in the form of a neural network could work even better.

• Clustering found other kinds of anomalies in the form of rare points and clusters.

• We extended this to ship data without much trouble.

• Speedups with 28 processors were 2-80 times, and speedups with a supercomputer were around 2000.

• It is clear that this task is well suited for distributed processing.
Collecting attack intelligence with honeypots
Types of honeypots we have run

• Honeypots are network nodes that have no purpose except to provide attack information.
  • Secure-shell (SSH) honeypots: These provide a wide range of attack information.
  • Web honeypots: These simulate a Web server and collect information about attempts to subvert it.
  • Cyber-physical system honeypots: These simulate utilities like power plants and collect information about attempts to control them.

• We collect data with log files and intrusion-detection systems.
• We use machine learning to find new types of cyberattacks.
• We will use deceptions to elicit more variety of responses.
Example data from the alert log for this attack

2007-01-30 20:52:33.506-08 | 2003 | 2 | MS-SQL Worm propagation attempt | 8 | misc-attack | 222.216.222.18 | 63.205.26.69 | 109 | 17 | 936746
2007-01-30 20:53:53.129-08 | 384 | 3 | ICMP PING | 4 | misc-activity | 72.54.246.133 | 63.205.26.87 | 113 | 1 | 936748
2007-01-30 20:53:55.336-08 | 384 | 3 | ICMP PING | 4 | misc-activity | 72.54.246.133 | 63.205.26.87 | 113 | 1 | 936749
2007-01-30 20:54:16.335-08 | 384 | 3 | ICMP PING | 4 | misc-activity | 72.54.246.133 | 63.205.26.78 | 113 | 1 | 936751
2007-01-30 21:01:49.934-08 | 491 | 2 | INFO FTP Bad login | 7 | bad-unknown | 63.205.26.69 | 61.247.253.72 | 127 | 6 | 936753
2007-01-30 21:02:02.945-08 | 491 | 2 | INFO FTP Bad login | 7 | bad-unknown | 63.205.26.69 | 61.247.253.72 | 127 | 6 | 936755
Example same-request reconnaissance of a local-area network

2007-01-29 20:10:56.238-08 | ICMP Destination Unreachable Host Unreachable | 4 | misc-activity | 63.211.230.114 | 63.205.26.69
2007-01-29 20:10:56.238-08 | ICMP Destination Unreachable Host Unreachable | 4 | misc-activity | 63.211.230.114 | 63.205.26.69
2007-01-29 20:10:57.253-08 | ICMP Destination Unreachable Port Unreachable | 4 | misc-activity | 63.205.51.6 | 63.205.26.69
2007-01-29 20:10:57.257-08 | ICMP Destination Unreachable Port Unreachable | 4 | misc-activity | 63.205.51.3 | 63.205.26.69
2007-01-29 20:10:57.266-08 | ICMP Destination Unreachable Port Unreachable | 4 | misc-activity | 63.205.51.4 | 63.205.26.69
2007-01-29 20:10:57.271-08 | ICMP Destination Unreachable Port Unreachable | 4 | misc-activity | 63.205.51.5 | 63.205.26.69
Example attack data, page 2 ("drill down" on day)

<table>
<thead>
<tr>
<th>Snort Alert Daily Totals, Jan 29 – Feb 4</th>
<th>312</th>
<th>329</th>
<th>11538</th>
<th>18602</th>
<th>17028</th>
<th>46437</th>
<th>339</th>
<th>312</th>
<th>226</th>
<th>226</th>
<th>435</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAD-TRAFFIC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ICMP</td>
<td>279</td>
<td>289</td>
<td>11466</td>
<td>15618</td>
<td>16727</td>
<td>46397</td>
<td>282</td>
<td>253</td>
<td>198</td>
<td>206</td>
<td>395</td>
</tr>
<tr>
<td>INFO</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2952</td>
<td>245</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MS-SQL</td>
<td>28</td>
<td>31</td>
<td>40</td>
<td>26</td>
<td>36</td>
<td>19</td>
<td>20</td>
<td>24</td>
<td>22</td>
<td>16</td>
<td>25</td>
</tr>
<tr>
<td>NETBIOS</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td>10</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>SCAN</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SHELLCODE</td>
<td>1</td>
<td>2</td>
<td>29</td>
<td>1</td>
<td>13</td>
<td>16</td>
<td>26</td>
<td>23</td>
<td>3</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>WEB-IIS</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WEB-PHP</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Example attack data, page 3 (drill down on alert)

<table>
<thead>
<tr>
<th>ICMP-Destination-Unreachable-Host-Unreachable</th>
<th>1</th>
<th>0</th>
<th>6458</th>
<th>11694</th>
<th>9438</th>
<th>31381</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICMP-Destination-Unreachable-Network-Unreachable</td>
<td>0</td>
<td>0</td>
<td>62</td>
<td>225</td>
<td>129</td>
<td>5472</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ICMP-Destination-Unreachable-Port-Unreachable</td>
<td>0</td>
<td>10</td>
<td>4567</td>
<td>3369</td>
<td>3654</td>
<td>6092</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>ICMP-Destination-Unreachable-Protocol-Unreachable</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>31</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ICMP-Echo-Reply</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ICMP-PING-CyberKit-2.2-Windows</td>
<td>33</td>
<td>25</td>
<td>15</td>
<td>17</td>
<td>24</td>
<td>15</td>
<td>25</td>
<td>16</td>
<td>15</td>
<td>10</td>
<td>33</td>
</tr>
<tr>
<td>ICMP-PING-Sun-Solaris</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ICMP-PING-Windows</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ICMP-PING</td>
<td>244</td>
<td>254</td>
<td>320</td>
<td>278</td>
<td>372</td>
<td>241</td>
<td>257</td>
<td>229</td>
<td>183</td>
<td>186</td>
<td>355</td>
</tr>
<tr>
<td>ICMP-Time-To-Live-Exceeded-in-Transit</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ICMP-redirect-host</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>15</td>
<td>3070</td>
<td>1707</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ICMP-redirect-net</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td>16</td>
<td>31</td>
<td>1422</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ICMP-traceroute</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>INFO-FTP-Bad-login</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2952</td>
<td>245</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>INFO-web-bug-1x-1-gif-attempt</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

2/28/2019
Count of Snort alerts in a year on a honeypot

Green represents attack variety, and blue represents attack volume.
Web-server attackers tell us what they want

<table>
<thead>
<tr>
<th>Web-Request Argument</th>
<th>Count</th>
<th>Email Name</th>
<th>User</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>717</td>
<td>test</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>492</td>
<td>info</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>admin</td>
<td>98</td>
<td>admin</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>test</td>
<td>61</td>
<td>sales</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>info</td>
<td>41</td>
<td>web</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>david</td>
<td>18</td>
<td>contact</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>michael</td>
<td>18</td>
<td>postmaster</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>mike</td>
<td>15</td>
<td>office</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>richard</td>
<td>15</td>
<td>spam</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>
Experiments we ran with a different deception every week

<table>
<thead>
<tr>
<th>Week</th>
<th>Number of packets</th>
<th>Number of alerts</th>
<th>Different alerts</th>
<th>ICMP alerts</th>
<th>TCP alerts</th>
<th>UDP alerts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>438661</td>
<td>388</td>
<td>4</td>
<td>388</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1191410</td>
<td>8589</td>
<td>24</td>
<td>8366</td>
<td>2185</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>1313693</td>
<td>259776</td>
<td>36</td>
<td>255744</td>
<td>4016</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>701771</td>
<td>2525</td>
<td>12</td>
<td>1940</td>
<td>584</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>906893</td>
<td>2823</td>
<td>17</td>
<td>2176</td>
<td>647</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>740769</td>
<td>6686</td>
<td>11</td>
<td>2990</td>
<td>3696</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>897552</td>
<td>3386</td>
<td>14</td>
<td>2144</td>
<td>1242</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>951556</td>
<td>2957</td>
<td>19</td>
<td>2651</td>
<td>306</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>618723</td>
<td>1325</td>
<td>13</td>
<td>757</td>
<td>568</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>541740</td>
<td>756</td>
<td>16</td>
<td>476</td>
<td>270</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>995235</td>
<td>2526</td>
<td>10</td>
<td>2270</td>
<td>256</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>807712</td>
<td>3711</td>
<td>15</td>
<td>3445</td>
<td>266</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>518659</td>
<td>488</td>
<td>5</td>
<td>488</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>1066743</td>
<td>4694</td>
<td>14</td>
<td>3082</td>
<td>1612</td>
<td>0</td>
</tr>
</tbody>
</table>
Tracking people with physical-motion analysis
Tracking technology for physical-motion analysis

• Automated surveillance technology has made progress.
• Many educational skills have a physical-motion component: physical education, theater, vocational training, working in groups.
• Good instructors can watch and give feedback – but events often happen quickly; video takes time to view.
• Cameras are usual for surveillance, but to save money, motion detectors and sonar suffice.
• Sensors can thoroughly monitor an area 10 meters by 10 meters for around $1000 US.
• We studied two Marine training tasks and the problem of detection suspicious terrorist activity in a public area.
Our two military testbeds

Marine patrol

Marine cordon and search
Technology for indoors tests

Most from www.phidgets.com:

- A collection of third-party sensors
- Interface through a USB port to a computer
- Software for data collection
Tracking Marines with sensors

• Use motion sensors to tell when Marines are in a room.
• Use sonar, narrow-infrared, and pressure strips to tell when they are near a door.
• Use light sensor to approximately locate them in a room using maps of light intensity.
• Use vibration sensor to tell when Marines are moving.
• Use orientation sensors to tell where they point their weapons.
• Use microphones and audio processing to recognize footsteps, door openings, and speech
### Exercise 14: Redundancy (Incorrect)

**Ahren’s Room Search (w/o grenade, door open)**

<table>
<thead>
<tr>
<th>M1 actions</th>
<th>M2 actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. M1 begins moving down the hall and DOES NOT report that he sees an open door as he’s approaching Rm. E120 (Ahren’s room).</td>
<td>1. M2 follows 5 feet behind M1.</td>
</tr>
<tr>
<td>2. M1 enters the room immediately, w/o a signal from M2, and begins searching with his weapon is at the ready.</td>
<td>2. M2 simply stands in the doorway while M1 searches Rm. E120.</td>
</tr>
</tbody>
</table>
| 3. a) M1 walks out of the room without warning, and follows M2 to the classroom door.  
b) M1 realizes he forgot to check inside the fridge in Ahren’s room, and informs M2 to pull security while he returns to check the fridge.  
c) M1 moves back out to the hallway once finished searching.  
d) M1 and M2 do not assume the stacked position and move immediately into the classroom. | 3. a) M2 walks from the open doorway to the entrance of the classroom.  
b) M2 DOES NOT pull security while M1 searches Ahren’s room for a second time.  
c) M1 and M2 do not assume the stacked position and move immediately into the classroom. |

**Classroom Search (door closed, Booby Traps Encountered)**

<table>
<thead>
<tr>
<th>M1 actions</th>
<th>M2 actions</th>
</tr>
</thead>
</table>
| 4. a) M1 follows 5 feet behind M2, and points his weapon in the same direction as M2’s weapon.  
b) M1 says nothing to the civilian.  
b) M1 does not provide security while M2 picks up the booby trap. | 4. a) M2 opens the door and moves into the classroom. He does not say anything to the civilian until he has cleared the south part of the room.  
b) Once M2 discovers the booby trap between C2 and C6, he simply picks the booby trap up and takes it with him. |
| 5. M1 follows behind M2 and helps to search the civilian.                  | 5. M2 then moves to search the civilian casually.                           |
| 6. a) M1, M2, and Civilian meet at the exit of the room. M1 and M2 escort the civilian to the hallway exit, | 6. a) M1, M2, and Civilian meet at the exit of the room. M1 and M2 escort the civilian to the hallway exit, |
Performance metrics and issues

Example metrics:
• How many of the potential threats they covered with their weapons
• How much of the rooms they searched
• How quickly they performed the task

Example issues:
• Did the Marines enter doors properly?
• Did they reconnoiter each room in the best order?
• Did the Marines point weapons accidentally at one another?
“Interesting events” for indoors training (loc. c2)

170-270 and 620-720 had correct execution; 400-480 and 810-890 did not.
2=noises, 9=talking, 12=infrared, 14=motion, 15=light, 17=sonar, 18=force.
Weapons coverage with the orientation sensors

\[ D = 0.5 \left(1 + \frac{1}{\pi}\right) \min(|\theta_1 - \theta_2|, 2\pi - |\theta_1 - \theta_2|) \left(|\cos^3 \phi_1| + |\cos^3 \phi_2|\right) \]

\( \theta \) is azimuth (bearing) angle, \( \phi \) is inclination angle from horizontal
Marine performance evaluation

Aggregates over exercises, squads, behaviors, and individuals

Metrics per timestep

Issues (problems) observed

Customization parameters

Marine locations

Marine orientations
Marine performance metrics

- Dispersion
- Collinearity
- Number of clusters
- Non Marine interactions
- Danger
- 360 awareness
- Weapons coverage
- Mobility
- Speed
- Safety of weapons
- Too close to window or door
- Too far from window or door
- Danger is surrounded
- Leader centrality
Yaw (azimuth) angle for a Marine in exercise 492
Exercise 492: 0 = black 3936 = red 3928 = green 3930 = magenta 3931 = blue 3934 = cyan 3935 = yellow

Time plot of individual issues
Time plot of group issues
All issues of one individual

Marine 3930 in exercise 492: dispersion problem red, windows/doors problems blue, flagging yellow, dangerous magenta, not pieing cyan
Detecting suspicious public behavior
Conclusions on physical-motion analysis

• Surveillance technology is a new tool for helping instructors assess student performance in physical activities.
• Video and virtual-reality playbacks of training can be usefully supplemented with statistics-based visualizations.
• Several kinds of visualizations are needed, each providing different insights into the data.
• Surveillance hardware is getting cheaper, and Kinect game tracking technology may be the future.