In 1969, the **Zodiac serial killer** sent two ciphers to San Francisco newspapers. The first cipher was solved very quickly by a high school teacher and his wife. Forty years later, the second cipher remains unsolved.

Some have attempted to decipher the 340-character cipher using n-gram frequency analysis. Problems:
- Ciphertext must be sufficiently large
- n-gram frequencies vary depending on source text

We use a dictionary-based attack:
Place words in one section of cipher to impose constraints on other sections of cipher. Correct placements produce partial word decodings in other sections of cipher text. Dictionary is indexed by constraints for fast lookup.

We search using frequency analysis zodiacdecoder project - performs GA search using frequency analysis http://code.google.com/p/zodiacdecoder/

**Evolutionary approach:**
- **Genome encodes attacks** as (word, position) tuples. Words are drawn from a fixed dictionary.
- Infeasible encodings are immediately rejected.
- Each individual starts with a single tuple. Subsequent generations add more tuples via crossover and mutation.
- Fitness function measures counts and coverage of potential words that appear when the genome’s words are plugged into the cipher text.

Example genome: 

```
((0,0), (1,7), (2,11))
```

- \(W_0\) (“killing”) at position 0
- \(W_1\) (“wild”) at position 7
- \(W_2\) (“game”) at position 11

Results: For dictionary sizes up to 1600 words, algorithm was able to find correct decodings for the 408-character cipher. 340-character cipher remains unsolved. We are working on making our technique more robust for future attacks.

<table>
<thead>
<tr>
<th>#Words</th>
<th>Correctness</th>
<th>(F_r)</th>
<th>(F_s)</th>
<th>Generations</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>1.0</td>
<td>50.000</td>
<td>50.000</td>
<td>874</td>
</tr>
<tr>
<td>850</td>
<td>1.0</td>
<td>82.074</td>
<td>82.074</td>
<td>698</td>
</tr>
<tr>
<td>1000</td>
<td>1.0</td>
<td>85.109</td>
<td>85.109</td>
<td>807</td>
</tr>
<tr>
<td>1285</td>
<td>1.0</td>
<td>104.170</td>
<td>104.170</td>
<td>1750</td>
</tr>
<tr>
<td>1600</td>
<td>0.9</td>
<td>110.130</td>
<td>120.984</td>
<td>3222</td>
</tr>
</tbody>
</table>

**Table 1:** Results of experimental runs for different word pool sizes. \(F_r\) is the evolved solution’s multi-objective fitness, and \(F_s\) is the multi-objective fitness of the known correct solution.