WonderWall
A Machine-Learned Network Filtering System

Abstract
The WonderWall system identifies normal and malicious network packets using a machine learning classifier trained on a research dataset. A virtualized software-defined network demonstrates the filtering engine at work, handling the traffic based on the classification labels.

Motivation & Goals
- Rules-based IDS such as Snort are rigid/rules-based
- Generating threat signatures, responding to alerts requires a lot of time and effort from network admins
- Also inaccurate. Snort classification statistics:
  - Precision: 68%
  - Recall: 23%
  - F-1 Score: 23%
- A network admin cannot blindly rely on filtering due to Snort rules alone. Many attacks are missed entirely!

Design & Implementation
- WonderWall is an attached network appliance
- High-level architecture: feedback loop
  - Backbone routers mirror protected network segment's ingress traffic to WonderWall
  - To minimize network latency, WonderWall appliance performs classification asynchronously
- When a threat is identified, offending traffic is blocked through OpenFlow "flow routing" rules
- Software-defined network management
  - Protected network segment's edge switches managed via OpenFlow-v1.5 protocol
- WonderWall itself is an OpenFlow controller written in Ryu (Python SDN framework)
- For development and testing, Mininet is used as a lightweight network simulation tool

Data
Data from UNB ISCX Intrusion Detection Evaluation set.
- Realistic, up-to-date and diverse packet captures.
- 13 Features (in order of importance)
  - Source Port, Destination, Source Bytes, Source Destination Bytes, Source TCP Flags, ...

Classification Performance
<table>
<thead>
<tr>
<th>Training Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>~1,500,000</td>
</tr>
<tr>
<td>Precision</td>
<td>97%</td>
</tr>
<tr>
<td>Recall</td>
<td>96%</td>
</tr>
<tr>
<td>F-1 Score</td>
<td>97%</td>
</tr>
</tbody>
</table>

Figure 2: Learning curve of decision tree classifier accuracy (training and cross-validated accuracies shown)

Conclusion
WonderWall is an alternative to traditional rules-based intrusion detection/prevention systems like Snort. Our machine learning approach to network filtering enables WonderWall to learn about new types of attacks with minimal effort required from a network administrator. In an industrial setting, a network capture dataset such as the UNB one we used could be distributed with WonderWall in order to establish a baseline of threat knowledge for fresh deployments of WonderWall.

Future Work
- Conduct a network performance impact study
- Create tooling to enable easy continuous learning on additional deployment-specific traffic captures and feedback from network administrator
- Distribute WonderWall controller across multiple nodes for improved fault-tolerance