SECURITY NEWS THIS WEEK

- Verizon Hit by Another Amazon S3 Leak

- researchers found a trove of sensitive corporate data in a publicly accessible Amazon S3 bucket

- 100MB files named ‘VZ Confidential’ and ‘Verizon Confidential’, some of which contained usernames, passwords and these credentials could have easily allowed access to other parts of Verizon's internal network
SECURITY NEWS THIS WEEK

- Verizon Hit by Another Amazon S3 Leak
  - the bucket appears to have been self-owned by a Verizon Wireless engineer and so wasn’t managed by the company (improperly configured S3)
  - Verizon: took the database offline soon after being informed
  - involving personal and account data on as many as six million Verizon customers
SECURITY NEWS THIS WEEK

- cloud-based or locally-hosted? Where is More Secure?
SECURITY NEWS THIS WEEK

- Global Cost of Cybercrime Soars 23% in a Year, from Accenture.

**FIGURE 1**
The global average cost of cyber crime over five years
US dollars

Legend
- Total average cost
- Five-year average

Percentage change in average cost over five years is 62 percent
Global Cost of Cybercrime Soars 23% in a Year, from Accenture.
SECURITY NEWS THIS WEEK

- Malware ($2.4m) and web attacks ($2m) are the most costly.

- Malicious insiders caused the most havoc, with related incidents taking on average 50 days to resolve, while ransomware attacks take over 23 days.
SECURITY NEWS THIS WEEK

- most efficient solutions

**FIGURE 20**
Cost savings when deploying enabling technologies

Legend
Consolidated view
n = 254 companies

- Annual cost savings (US$)

<table>
<thead>
<tr>
<th>Solution</th>
<th>Cost Savings (US$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Security intelligence systems</td>
<td>$2,839,707</td>
</tr>
<tr>
<td>Advanced identity and access governance</td>
<td>$2,362,009</td>
</tr>
<tr>
<td>Automation, orchestration and machine learning</td>
<td>$2,178,378</td>
</tr>
<tr>
<td>Extensive use of cyber analytics and User Behavior Analytics</td>
<td>$1,665,900</td>
</tr>
<tr>
<td>Advanced perimeter controls</td>
<td>$1,017,932</td>
</tr>
<tr>
<td>Extensive deployment of encryption technologies</td>
<td>$997,062</td>
</tr>
<tr>
<td>Extensive use of data loss prevention</td>
<td>$983,900</td>
</tr>
<tr>
<td>Enterprise deployment of Governance, Risk &amp; Compliance</td>
<td>$877,211</td>
</tr>
<tr>
<td>Automated policy management</td>
<td>$590,342</td>
</tr>
</tbody>
</table>
SECURITY NEWS THIS WEEK

- NVIDIA GTC (GPU Technology Conference)

- Jen-Hsun Huang: GPU is Moore's Law in the era of artificial intelligence
SECURITY NEWS THIS WEEK

- Current Improvement: huge purchase and operating cost gap
- Traditional solution to get a 45,000 photos per second processing capacity:
  - four racks, 160 CPU servers, at 65 kilowatts of power
- if you replace NVIDIA's products (45000 photos per second):
  - only 1 NVIDIA HGX server with 8 V100 GPU, at 3 kilowatts
Q: Which one do you think is more important in the future artificial intelligence, GPU or FPGA?

A: The difference is huge. FPGA is very flexible, and you can even use it in the network card. However, the GPU is not so that flexible, it is only a parallel computing accelerator. In order to take full advantage of the GPU, we made the decision to make our GPU Tensor executive processor three years ago. Now it has become the world's best Tensor processor.

Q: Are you (NVIDIA) going to do research on autopilot algorithms?

A: Absolutely! Otherwise, you do not know what the processor is going to deal with and what kind of processor is more appropriate.
AI Chips

- Google: TPU with 10x processing capacity on Tensorflow
- Baidu & Xilinx: XPU

Jen-Hsun Huang:

- TPU and other products only support their own company's computing framework, like the Dumb-phone.
- GPU has more functions, broader application scene, richer ecologically, like the smart phone.
SECURITY NEWS THIS WEEK

- A joke about iPhone X face detection on a forum

- How to evaluate the face detection technique?

  - One day, your girl friend come to you with your phone, saying “can you have a look at the online store? I’m not sure which dress is better for me.” You do have a look at the screen.“Bi~, paid successfully!”

  - Another day, your girl friend come to you again with your phone, saying “Your ex sent you an intimate message! I will not forgive you if you can not explain well today!” And you’re amazing, “How can it be! let me see!” “Bi~, your dress paid successfully!”
POISONING ATTACK AGAINST SUPPORT VECTOR MACHINE

CS6604 Shengzhe Xu
CONTENTS

- Key Concepts
- Attack Algorithm
- Implementation and Experiments
- Compare to the Poisoning Attack Against Neural Network, 2017
- Conclusion and Discussion
KEY CONCEPTS IN THIS PAPER
KEY CONCEPT - SVM & KERNEL

\[ \overrightarrow{w}^T \overrightarrow{x} + b \]

\[ \overrightarrow{w}^T \phi(\overrightarrow{x}) + b \]

Feature Map
KEY CONCEPTS

- hinge loss

- an objective function (or loss function): \( l(y) = \max(0, 1 - t \cdot y) \)

- \( y \) is the predicted value and \( t \) is the target value

\[
\max_{x_c} L(x_c) = \sum_{k=1}^{m} (1 - y_k f_{x_c}(x_k))^+ = \sum_{k=1}^{m} (-g_k)^+ \quad (1)
\]

- dataset: Dtr, Dval
ATTACK ALGORITHM IDEA
ALGORITHM IDEA

- poisoning attack algorithm

- target: to find a point \((x_c, y_c)\), whose addition to \(D_{tr}\) maximally decrease the SVM’s classification accuracy.

- quantization: to maximize the hinge loss incurred on \(D_{val}\) by the SVM trained on \(D_{tr} \cup (x_c, y_c)\)

- gradient used for optimizing the attack (gradient ascent)

\[
\frac{\partial L}{\partial u} = \sum_{k=1}^{m} \left\{ M_k \frac{\partial Q_{sc}}{\partial u} + \frac{\partial Q_{kc}}{\partial u} \right\} \alpha_c, \quad (10)
\]

\[
M_k = -\frac{1}{\zeta} \left( Q_{ks} (\zeta Q_{ss}^{-1} - uu^T) + y_k v^T \right)
\]
ALGORITHM IDEA

- poisoning attack algorithm
  - a random point of the non-attack class is selected and its label is flipped to serve as the starting point.
  - a gradient ascent method is then used to refine this attack until its termination condition is satisfied (step change less than a threshold).
  - seems like a reverse operation to training a SVM
IMPLEMENTATION & EXPERIMENTS ON ARTIFICIAL DATA
EXPERIMENTS ON ARTIFICIAL DATA

- data source: a two-dimensional data generation model in which each class follows a Gaussian distribution.
- \( x \) in \([-5, 5]^2\)
- kernel tested: both linear and RBF
poisoning attack algorithm
artificial data - linear kernel

Algorithm 1 Poisoning attack against SVM

Input: $\mathcal{D}_{tr}$, the training data; $\mathcal{D}_{val}$, the validation data; $y_c$, the class label of the attack point; $x_c^{(0)}$, the initial attack point; $t$, the step size.

Output: $x_c$, the final attack point.

1: $\left\{\alpha_i, b\right\} \leftarrow$ learn an SVM on $\mathcal{D}_{tr}$.
2: $k \leftarrow 0$.
3: repeat
4: Re-compute the SVM solution on $\mathcal{D}_{tr} \cup \{x_c^{(p)}, y_c\}$ using incremental SVM (e.g., Cauwenberghs & Poggio, 2001). This step requires $\{\alpha_i, b\}$.
5: Compute $\frac{\partial L}{\partial u}$ on $\mathcal{D}_{val}$ according to Eq. (10).
6: Set $u$ to a unit vector aligned with $\frac{\partial L}{\partial u}$.
7: $k \leftarrow k + 1$ and $x_c^{(p)} \leftarrow x_c^{(p-1)} + tu$
8: until $L(x_c^{(p)}) - L(x_c^{(p-1)}) < \epsilon$
9: return: $x_c = x_c^{(p)}$

$$\frac{dO_{kc}}{dx_c} = \frac{d}{dx_c} y_k y_c K(x_k, x_c) = y_k y_c \cdot x_k$$
poisoning attack algorithm
artificial data - RBF kernel

Algorithm 1 Poisoning attack against SVM

Input: $D_{tr}$, the training data; $D_{val}$, the validation data; $y_c$, the class label of the attack point; $x_c^{(0)}$, the initial attack point; $t$, the step size.

Output: $x_c$, the final attack point.

1: $\{\alpha_i, b\} \leftarrow$ learn an SVM on $D_{tr}$.
2: $k \leftarrow 0$.
3: repeat
4: Re-compute the SVM solution on $D_{tr} \cup \{x_c^{(p)}, y_c\}$ using incremental SVM (e.g., Cauwenberghs & Poggio, 2001). This step requires $\{\alpha_i, b\}$.
5: Compute $\frac{\partial L}{\partial u}$ on $D_{val}$ according to Eq. (10).
6: Set $u$ to a unit vector aligned with $\frac{\partial L}{\partial u}$.
7: $k \leftarrow k + 1$ and $x_c^{(p)} \leftarrow x_c^{(p-1)} + tu$
8: until $L(x_c^{(p)}) - L(x_c^{(p-1)}) < \epsilon$
9: return: $x_c = x_c^{(p)}$

$$\frac{dQ_{kc}}{dx_c} = y_k y_c \cdot K(x_k, x_c) \cdot \gamma \cdot (x_k - x_c)$$
EXPERIMENTS ON REAL DATA

- data source: MNIST, handwritten digit classification task
- kernel tested: only linear
- preprocess:
  - number of features: \( d = 28 \times 28 = 784 \)
  - normalized each feature from \([0, 255]\) to \([0, 1]\)
  - size of Dtr is 100 and size of Dval is 500
  - size of Dts is about 2000 per digit
EXPERIMENTS RESULTS
EXPERIMENTS RESULTS

classification error (7 vs 1)

- validation error
- testing error

classification error (4 vs 0)
CONCLUSIONS FOR SVM ATTACK

- Attack is not only works well on validation dataset but also on test datasets
- SVM may be very vulnerable to poisoning?
GENERATIVE POISONING ATTACK AGAINST NEURAL NETWORK (ESP. DNN), YANG 2017

- examine traditional gradient-based poisoning attack on NN and identify the poisoned data generation rate, the bottleneck of its implementation
- propose a generative method to substantially speed up the poisoned data generation rate with slightly degraded model attack effectiveness, i.e., target model accuracy degradation
- proposed a loss-based countermeasure to detect the poisoning attack with very minimum overhead
- evaluated the effectiveness by performing experiments on MNIST and CIFAR-10 datasets under different configurations
**POISONING ATTACK AGAINST NN**

Fig. 1: An overview of direct gradient method.

Fig. 2: An overview of the generative method.
POISONING ATTACK AGAINST NN

(a) Start from normal data “5” by applying the direct gradient method.

(b) Start from a uniform distribution sampling by applying the direct gradient method.

(c) Start from normal data “5” by applying the generative method.

(d) Start from normal data “bird” by applying the generative method.
DISCUSSIONS

- In the real data test, why they restrict: linear only and fixed $C=1$?
- Do you think it's just an ideal attack or an applied one for other scenarios or models?
- What if the ML model is not a white box for attackers?
- How to improve robustness to poisoning?
- What can be learned from the paper from today's point of view?
THANKS!