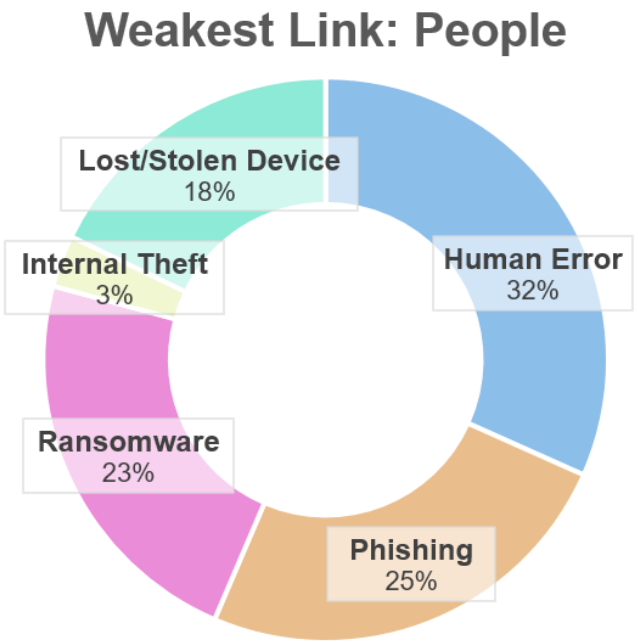
Advanced Countermeasures: Integrating SSDs Into Cyber-Security Defense

Sebastien Jean, Phison Electronics

Greg Scasny, Cigent Technology

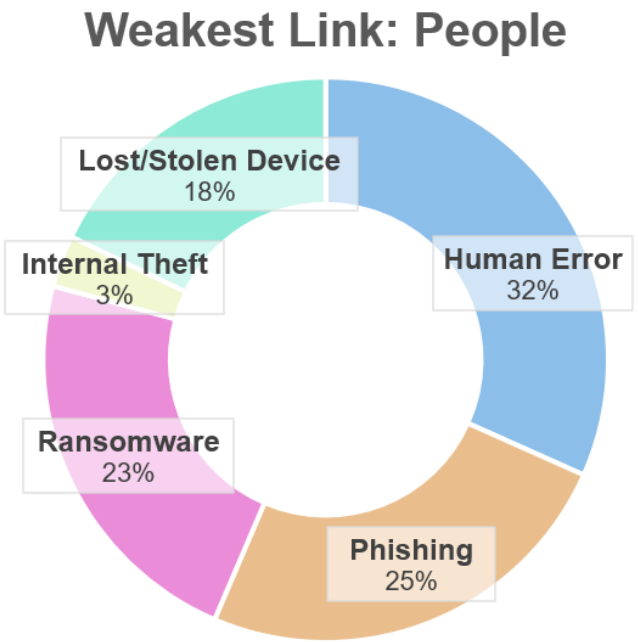
Cyber-security Landscape

- What's at risk
  - Data can have sentimental value (pictures)
  - Data can be very personal (medical records)
  - Data can be very expensive to acquire (DNA Sequencing)
  - Data can be critical for the operation of a business or even a city
- Problem
  - Data Theft and Data Ransoming are a growing problem
  - Bypassing standard antivirus tools is relatively easy if you understand how binaries are structured
  - The days of quick attacks are over
  - Attackers are taking their time to study the network, find the most valuable data or do the most possible damage by destroying backups
- Solution
  - AI/ML cyber-security suite that is tightly integrated with the storage align the protection system with the threat landscape



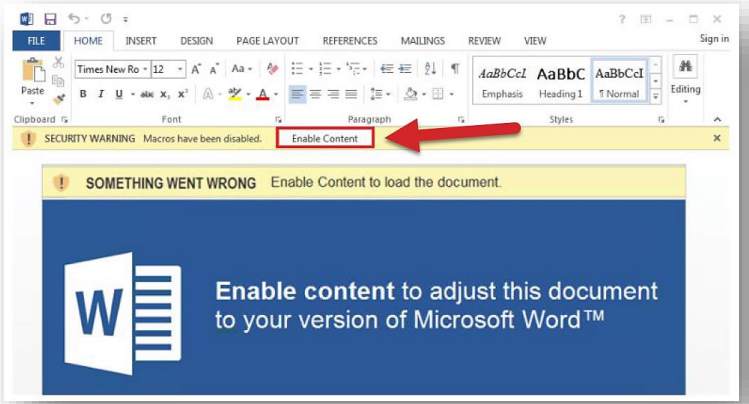
Cyber-security Landscape

- June/2019 – Florida Ransomware Headlines
  - Lake City: Pays $460,000 in Ransom to Cyber attackers
  - Key Biscayne: Someone clicks link, again, giving ransomware
  - Riviera Beach: Agrees to pay ransomware hackers $600,000 to unlock its data
- Current methods used to combat cyber security threads include: People, Process, Technology
  - Despite constant training, people remain the weakest link, accounting for 43% of all the security breaches
- How can AI/ML help detect and stop cyberattacks beyond what is being done today?
  - The Florida Ransomware attacks represent an advanced attack knows as a “triple threat”
  - What does this mean and how can we protect ourselves?



Getting into the network

- An unwitting victim receives a phishing email that usually contains a macro-enable office document
  - When “Enable Content” is pressed, a macro runs that spawns a command shell and downloads the “Emotet” trojan
  - Emotet was identified in 2014; it attempts to steal sensitive information
  - It has advanced polymorphic functionality that helps evade signature-based anti-malware products
  - It detects if it is in a VM/Sandbox and will remain dormant
  - Designed to spread to other computers on the network
  - US Department of Homeland Security concluded that Emotet is one of the most costly and destructive malwares, costing upward of $1M per incident to clean up



“It all starts with a single user - click”

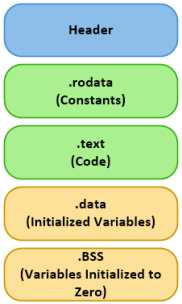
AI Detection & Prevention

- Reviewing attack vectors
  - Users are trained not to open attachments, but if the email appears to be internal, they assume it's safe
  - Office Macros are heavily used in corporate environments and while IT experts prefer to disable them, user backlash usually gets them re-enabled
  - Powershell is used in everything from auditing, account management, system hardening and configuration management
- These tools are entrenched, so why not use AI/ML to help detect unusual activity?
  - Attacks are typically structured as a chain, where one step allows the next step to continue
  - Breaking that chain will stop the attack
  - Machine Learning algorithms such as decision trees and nearest-neighbor type algorithms can detect deviations from baseline behavior
  - No need to search for specific file names or signatures
  - Enable maximum productivity by focusing on system activity that is outside the norm

Breaking Virus Scanner

- Polymorphic code or self-mutating sounds cool, but what does it mean?
  - The binary understands its own code structure
  - It can move sections around and self-edit, but still respects the rules of binary layout
  - Can be based on a valid executable that simply changes one of the branch calls to execute the attack function
  - Adds random symbols and instructions calls that are not actually accessed
  - SHA-256 hash will produce a substantially different signature even when only a few bytes are changed
  - Result is that the attacking exe passes the signature check
  - Though setting up this self-editing code is tedious, it is not that hard to do if you understand how compilers work

Basic Bin Layout



Defeating Virus Scanner

```
Contents of section .rodata:
 4005f8 01000200 57696E64 6F777300 00        ....Windows.
Contents of section .text:
 400410 31ed4989 d15e4889 e24883e4 f0505449  1.I..^H..H...PTI
 400420 c7c0a005 400048c7 c1100540 0048c7c7  ....@.H....@.H..
 400430 f4044000 e8c7ffff fff49090 4883ec08  ..@.........H...
```

This part isn't actually called by any branch in the bin. It's dead code.

```
Hash / Signature: a50087195cd390eececbcb7d4bbfef7b
```

Polymorphic == Self-edits to Bin

```
Contents of section .rodata:
 4005f8 01000200 57696E64 6F777300 00        ....Windows.
Contents of section .text:
 400410 31ed4989 d15e4889 e24883e4 f0505449  1.I..^H..H...PTI
 400420 c7c0a005 400048c7 c1100542 0049c7c7  ....@.H....B.I..
 400430 f4044000 e8c7ffff fff49090 4883ec08  ..@.........H...
```

We can put any hex value we want here

```
Hash / Signature: 0c2f38b0e02b36d29888d7638e6fdc40
```

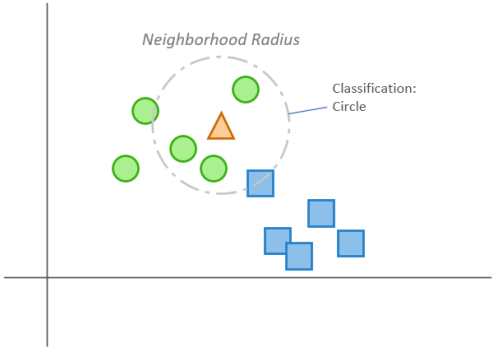
AI Detection & Prevention

- The AI/ML is trained to look for abnormal execution patterns
  - Normal usage of system binaries
  - Normal use (or non-use) of system commands such as “net”, “taskkil”, “psexe”, “vssadmin”, etc…
  - Normal lateral network communications
  - Normal access patterns as the encryption and removal of many files at once from storage

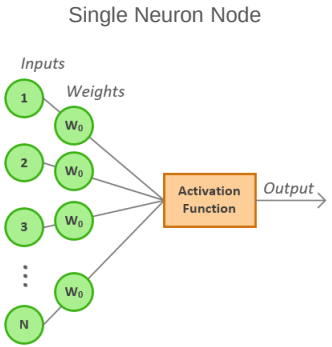
- Layered approach
  - Attackers use many different tools together to attack a system
  - The solution must take a similar approach
  - Sensors come with a pre-trained baseline and adapt based on the environment and specific user
  - Here are a few examples of Input Sensors:

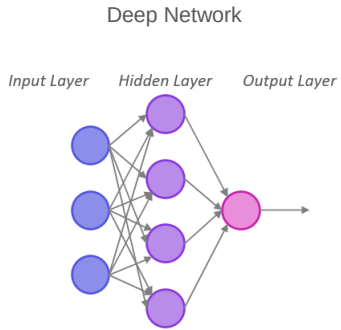
Training AI/ML Examples

- The K-Nearest Neighbor (KNN) algorithm executes very quickly and has good predictive power
  - In this example we have a set of circles and squares representing known items.
  - The axis represent sensor inputs, though there is no reason to limit to only two inputs (it's just easier to draw)
  - The objective is to classify the triangle based on how many (K) neighbors it has



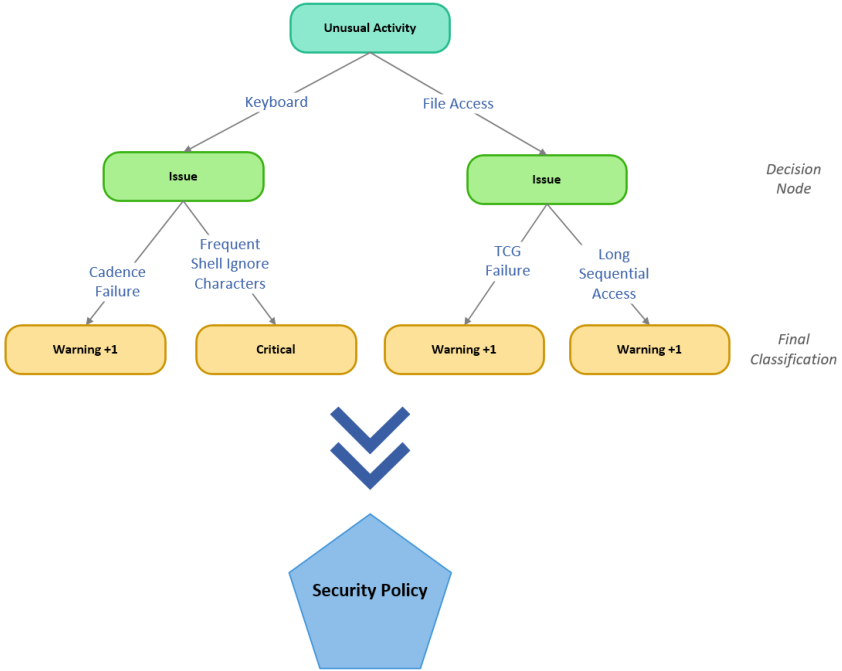
- A Neural Network (NN) can be trained dynamically
  - In this example we look at keyboard typing cadence and character usage
  - Input Network-1 is trained on character frequency
  - Input Network-2 is trained on delay from last keystroke
  - Hidden Layers add more complexity to the categorization
  - Output Node balances the input to make a determination





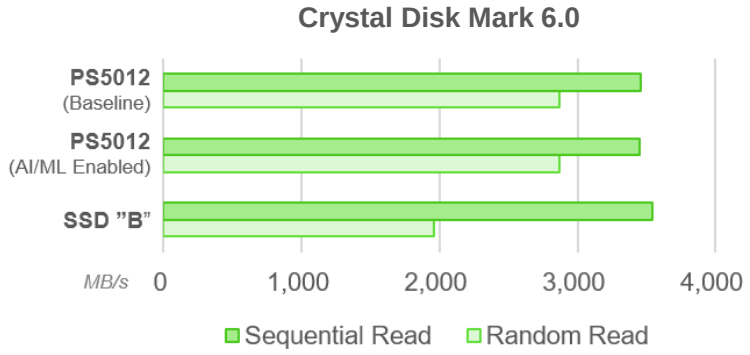
Training AI/ML Examples

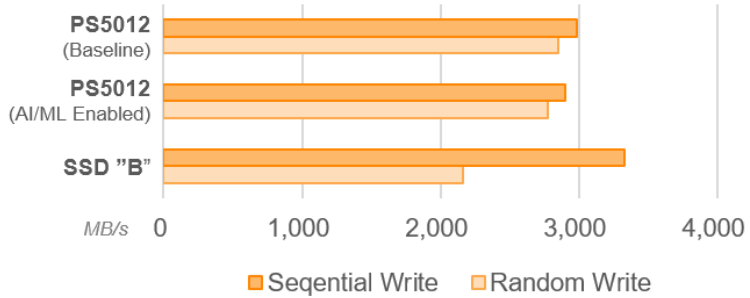
- A decision tree looks a lot like a flow chart
  - The training is a lot simpler as the data it works with does not need to be process or normalized
  - The Decision Tree works together with the KNN and NN based categorization engines to classify activity and then make a decision
  - The categories and issues are automatically identified by the learning algorithm based on the data collected
  - The objective with each split is to obtain many items of a specific class in that category
  - Apply regression function with a simple square of the error to prune out bad branches
  - Stop when minimum number of training issues assigned to each class is below a set target
  - The final classification is submitted to the security policy set up by the end-user or security officer



Sensor: SSD Storage

- The SSD can become another sensor input into the AI/ML engine
  - Based TCG Opal Self-Encrypting Drive (SED)
  - The SSD is divided into TCG Opal Ranges
  - The SSD maintains logs for each TCG Opal Ranges that can be analyzed by the AI/ML engine
- Secure SSD Sensor
  - Careful Hardware and Firmware design ensures no degradation in performance
  - Projected impact of logging on TBW is ~24B / 4K (0.1%)
  - No meaningful impact on drive lifespan
  - No significant degradation in performance





Key Takeaway

1. Data Theft and Data Ransoming are a growing problem
2. Attackers take their time to learn the network: Maximize Theft & Damage
3. Attackers can read white papers too, they will adopt AI as well
4. The only thing that beats AI is a better AI, simple rule-based logic is obsolete
5. Applying adaptive AI/ML gives a realistic chance of breaking the attack chain
6. Pushing the detection and response further down the data stack (all the way to the storage device) is a force multiplier for AI/ML defense
7. AI/ML cyber-security + tightly integrated storage = solid protection

PHISON

Knows What You Need

"PUSHING BOUNDARIES"

YOUR PCIe GEN-4 SSD LEADER

VISIT US AT BOOTH #219!