Machine Learning in Cyber Security

Definitions

- **Machine Learning (ML)**

Google’s definition - Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data.

Definitions (Contd.)

- **Data Analytics**

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making.

Definitions (Contd.)

- **Cyber Security**

Cybersecurity is the body of technologies, processes and practices designed to protect networks, computers, programs and data from attack, damage or unauthorized access. In a computing context, **security** includes both **cybersecurity**and physical **security**.

Definitions (Finally.)

- **ML + Data Analytics + Cyber Security**

Machine learning has been quickly adopted in cybersecurity for its potential to automate the detection and prevention of attacks, particularly for next-generation antivirus (NGAV) products. ML models in NGAV have fundamental advantages compared to traditional AV, including the higher likelihood of identifying novel, zero-day attacks and targeted malware, an increased difficulty of evasion, and continued efficacy during prolonged offline periods

An Overview of Machine Learning

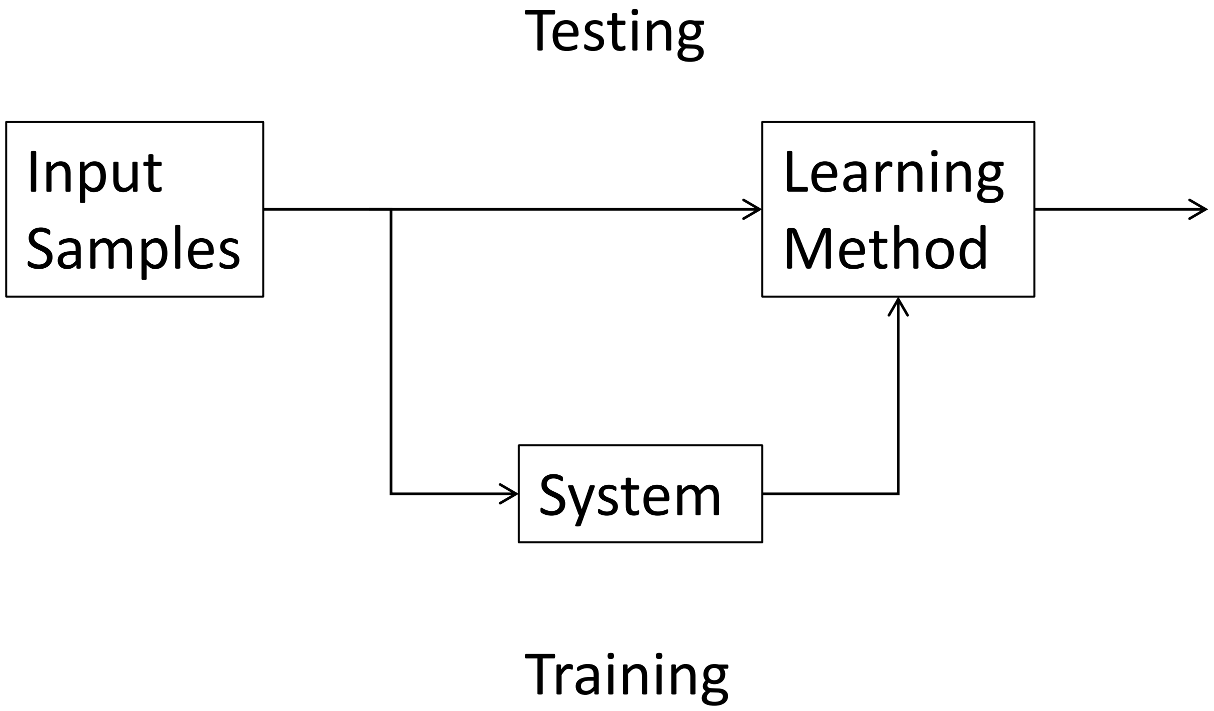
- What is machine learning?
- Learning system model
- Training and testing
- Performance
- Algorithms
- Machine learning structure
- Learning techniques
- Applications

What is machine learning?

- A branch of **artificial intelligence**, concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data.

- As intelligence requires knowledge, it is necessary for the computers to acquire knowledge.

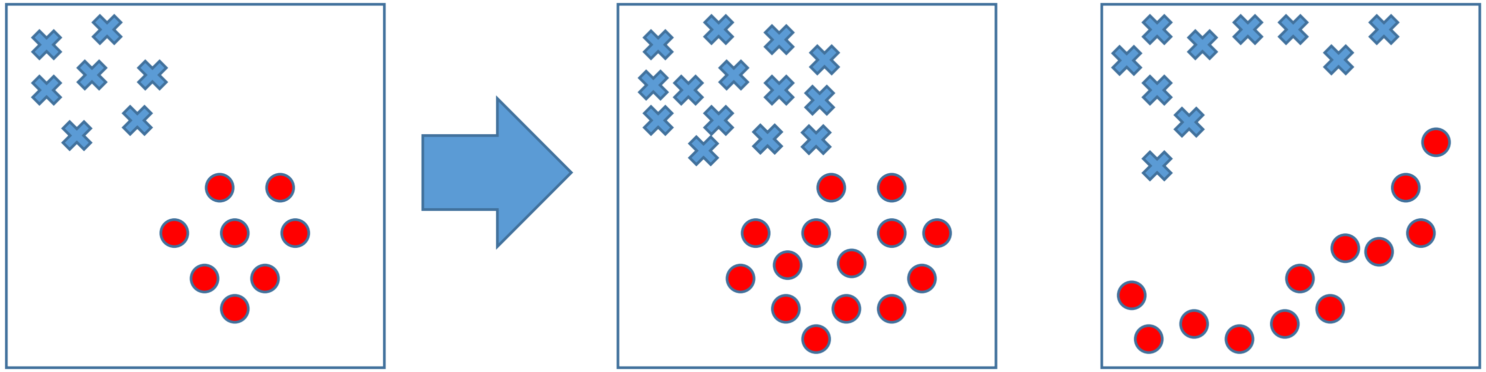
Learning system model



Training and testing

- Training is the process of making the system able to learn.

- No free lunch rule:
  - Training set and testing set come from the same distribution
  - Need to make some assumptions or bias



Performance

- There are several factors affecting the performance:
  - **Types of training** provided
  - The form and extent of any initial **background knowledge**
  - The **type of feedback** provided
  - The **learning algorithms** used

- Two important factors:
  - Modeling
  - Optimization

Algorithms

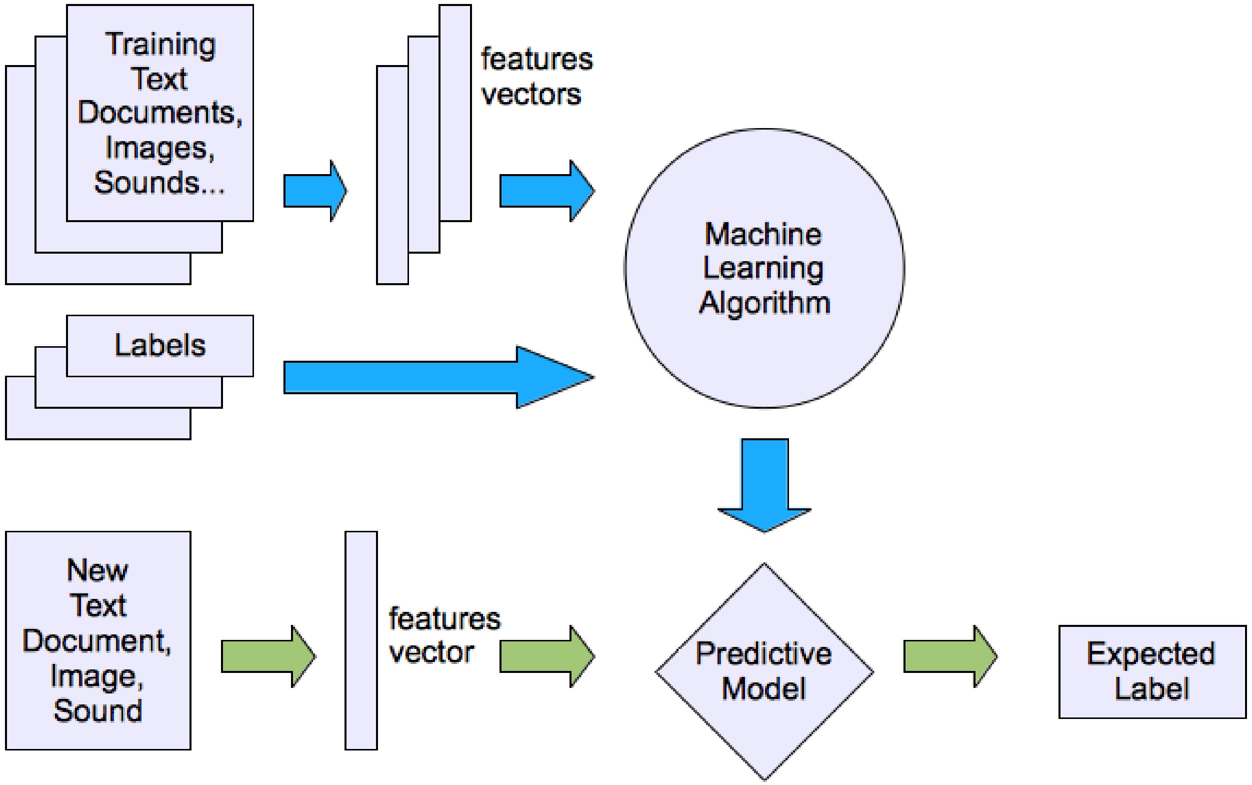
- The success of machine learning system also depends on the algorithms.
- The algorithms control the search to find and build the knowledge structures.
- The learning algorithms should extract useful information from training examples.

Algorithms

- **Supervised learning**
  - Prediction
  - Classification (discrete labels), Regression (real values)
- **Unsupervised learning**
  - Clustering
  - Probability distribution estimation
  - Finding association (in features)
  - Dimension reduction
- **Semi-supervised learning**
- **Reinforcement learning**
  - Decision making (robot, chess machine)

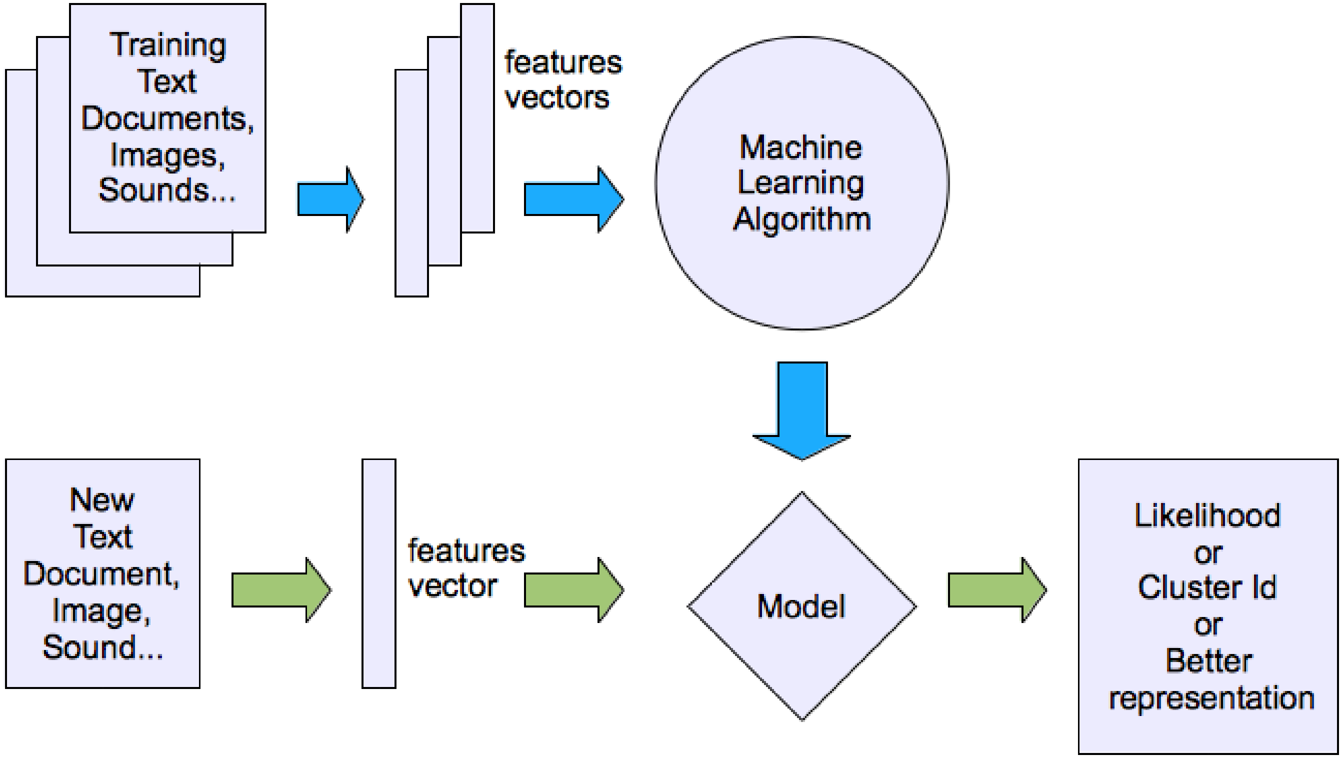
Machine learning structure

- Supervised learning


Training Text Documents, Images, Sounds...
features vectors
Labels
Machine Learning Algorithm
New Text Document, Image, Sound
features vector
Predictive Model
Expected Label

Machine learning structure

- Unsupervised learning


Training Text Documents, Images, Sounds...
features vectors
Machine Learning Algorithm
New Text Document, Image, Sound...
features vector
Model
Likelihood or Cluster Id or Better representation

Some Examples

- SPAM detection
    - Distinguish between SPAM and legitimate email
    - % of emails correctly classified
    - Hand-labeled emails
- Detecting catalog duplicates
    - Distinguish between duplicate and non-duplicate catalog entries
    - False positive/negative rate based on business criteria
    - *H*and-labeled duplicates and non-duplicates
- Go learner
    - *P*laying Go
    - % of games won in tournament
    - Practice games against itself

MALWARE

What is malware

Malware refers to malicious software perpetrators dispatch to infect individual computers or an entire organization's network. It exploits target system vulnerabilities, such as a bug in legitimate software (e.g., a browser or web application plugin) that can be hijacked.

A malware infiltration can be disastrous—consequences include data theft, extortion or the crippling of network systems.

Common malware types

There are numerous malware types, each having their own application area and focus. Seven of the most common variations are as follows:

Ransomware: Once installed, this malware encrypts files on a computer and/or across an extended network. A popup display informs the user that unless a ransom is paid, their files will remain encrypted.

Ransomware usually arrives as an email attachment or is unwittingly downloaded from a malicious website.

A new business model called ransomware as a service (RaaS) has recently appeared. Using it, amateur hackers (a.k.a., "script kiddies") license existing malware to execute a RaaS assault. In the event of success, a percentage of the ransom goes to the malware author.

MALWARE

Worms – These were originally designed to infect a computer, clone itself, and then infect additional computers via another medium, such as email.

Perpetrators use worms to create botnetsfrom a large numbers of compromised connected devices (e.g., mobile phones or PCs). Such devices are known as “zombies” because their owners are oblivious to the infection and that their systems are used as part of a much larger attack, such as a distributed denial of service (DDoS).

Worm examples include:

NgrBot– This worm propagates through chat messengers and social networking sites. Perpetrators use social engineering to encourage downloading of the malware that, once installed, turns the user’s machine into a zombie participating in a massive botnet. It also stops infected systems from being updated and can steal login credentials and other sensitive information.

ILOVEYOU– This has been deployed using a social engineering attack that encouraged people, through the enticement of a possible love interest, to open an email attachment containing the worm. A Visual Basic script is run that then overwrites various file types. The worm has infected an estimated 45 million computers.

Trojan – A Trojan appears legitimate but carries a dangerous payload. While it doesn’t replicate itself as do worms, it typically comes packaged with additional malware types—including backdoors, rootkits, ransomware and spyware.

The banking industry is a favorite target of Trojan attacks. For instance, the Tiny Banker Trojan (Tinba) malware, which is executed via the Rig exploit kit. Installation is achieved by first locating a software vulnerability on the target computer. It then overlays a spoofed screen requesting personal information, including credit card details, whenever the system user visits a bank site (see below).

MALWARE

Worms – These were originally designed to infect a computer, clone itself, and then infect additional computers via another medium, such as email.

Perpetrators use worms to create botnetsfrom a large numbers of compromised connected devices (e.g., mobile phones or PCs). Such devices are known as “zombies” because their owners are oblivious to the infection and that their systems are used as part of a much larger attack, such as a distributed denial of service (DDoS).

Worm examples include:

NgrBot– This worm propagates through chat messengers and social networking sites. Perpetrators use social engineering to encourage downloading of the malware that, once installed, turns the user’s machine into a zombie participating in a massive botnet. It also stops infected systems from being updated and can steal login credentials and other sensitive information.

ILOVEYOU– This has been deployed using a social engineering attack that encouraged people, through the enticement of a possible love interest, to open an email attachment containing the worm. A Visual Basic script is run that then overwrites various file types. The worm has infected an estimated 45 million computers.

Trojan – A Trojan appears legitimate but carries a dangerous payload. While it doesn’t replicate itself as do worms, it typically comes packaged with additional malware types—including backdoors, rootkits, ransomware and spyware.

The banking industry is a favorite target of Trojan attacks. For instance, the Tiny Banker Trojan (Tinba) malware, which is executed via the Rig exploit kit. Installation is achieved by first locating a software vulnerability on the target computer. It then overlays a spoofed screen requesting personal information, including credit card details, whenever the system user visits a bank site (see below).

MALWARE

Rootkits – These are a prepared, customizable software. They grant access to sensitive parts of an application, enable the execution of files and can even change system configurations.

Typically deployed through a social engineeringattack (e.g., phishing—resulting in the theft of a user's login credentials—its installation gains access to a network. The rootkit can then subvert any anti-malware software that might otherwise be able to detect it, giving the perpetrator free reign to install additional malware.

Examples of rootkits include Flame used in cyberespionage attacks to steal screenshots, record keystrokes and monitor network traffic. It was most notably used to disrupt Iranian oil refinery production in 2012.

Backdoors – A backdoor negates normal authentication required to access a system, such as via a webserver or database. Often its installation is part of a targeted assault; after researching a victim, social engineering is used to steal login credentials and gain access to an application.

Backdoors avoid detection and are used to set up a control center. This lets the perpetrator remotely update malware and initiate system commands.

Backdoors are used for many malicious activities, including data theft, denial of serviceassaults and infection of your visitors' computers. It's also an initial step when executing an advanced persistent threat(APT) assaults.

Backdoors have recently been found in a number of Internet of Things (IoT) devices, such as security Wi-Fi cameras used by organizations. Once an IoT device has been hacked and turned into a backdoor, it effectively provides a gateway into that network.

Adware – One of the earliest malware types, adware originated in the days of freeware. The software was free, but included popup ads that appeared whenever you used it. While annoying, it wasn't malicious.

Today your system can be infected from visiting a compromised website where its malware-laden adware, using a browser vulnerability, installs itself.

Spyware – This malware variant gathers personal data and sends it to a third-party without your knowledge or consent.

A highly malicious spyware type is a keylogger. Once installed, it tracks keyboard entries and sends the data, including login credentials, to the perpetrator.

MALWARE

Rootkits – These are a prepared, customizable software. They grant access to sensitive parts of an application, enable the execution of files and can even change system configurations.

Typically deployed through a social engineeringattack (e.g., phishing—resulting in the theft of a user's login credentials—its installation gains access to a network. The rootkit can then subvert any anti-malware software that might otherwise be able to detect it, giving the perpetrator free reign to install additional malware.

Examples of rootkits include Flame used in cyberespionage attacks to steal screenshots, record keystrokes and monitor network traffic. It was most notably used to disrupt Iranian oil refinery production in 2012.

Backdoors – A backdoor negates normal authentication required to access a system, such as via a webserver or database. Often its installation is part of a targeted assault; after researching a victim, social engineering is used to steal login credentials and gain access to an application.

Backdoors avoid detection and are used to set up a control center. This lets the perpetrator remotely update malware and initiate system commands.

Backdoors are used for many malicious activities, including data theft, denial of serviceassaults and infection of your visitors' computers. It's also an initial step when executing an advanced persistent threat(APT) assaults.

Backdoors have recently been found in a number of Internet of Things (IoT) devices, such as security Wi-Fi cameras used by organizations. Once an IoT device has been hacked and turned into a backdoor, it effectively provides a gateway into that network.

Adware – One of the earliest malware types, adware originated in the days of freeware. The software was free, but included popup ads that appeared whenever you used it. While annoying, it wasn't malicious.

Today your system can be infected from visiting a compromised website where its malware-laden adware, using a browser vulnerability, installs itself.

Spyware – This malware variant gathers personal data and sends it to a third-party without your knowledge or consent.

A highly malicious spyware type is a keylogger. Once installed, it tracks keyboard entries and sends the data, including login credentials, to the perpetrator.

MALWARE

Malware detection and removal

Imperva has a number services that prevent malware installation while weeding out existing infections on web application servers.

Web Application Firewall (WAF)–Deployed at the edge of your network, Imperva cloud PCI DSS compliant service uses signature, behavioral and reputational analysis to block all malware injection attacks on your websites and web applications. Imperva cloud WAF is offered as a managed service and maintained by a dedicated security team.

Backdoor Protect– A service that intercepts communication attempts with backdoor shells on your web server. By tracing these requests, the service is able to pinpoint the most highly obfuscated malware, even if it was installed on your web server long before you onboarded Imperva cloud security services.

Login Protect – A flexible two-factor authentication (2FA) solution that requires zero integration and can be instantly deployed on any Imperva cloud-protected URL address. The service prevents perpetrators from using stolen login credentials to obtain network access and install rootkits and backdoors on your web servers.

ANOMALY DETECTION

What is anomaly detection?

Anomaly detection is any process that finds the outliers of a dataset; those items that don’t belong. These anomalies might point to unusual network traffic, uncover a sensor on the fritz, or simply identify data for cleaning before analysis.

In today’s world of distributed systems, managing and monitoring the system’s performance is a chore—albeit a necessary chore. With hundreds or thousands of items to watch, anomaly detection can help point out where an error is occurring, enhancing root cause analysis and quickly getting tech support on the issue. Anomaly detection helps the monitoring cause of chaos engineeringby detecting outliers, and informing the responsible parties to act.

In enterprise IT, anomaly detection is commonly used for:

Data cleaning

Intrusion detection

Fraud detection

Systems health monitoring

Event detection in sensor networks

Ecosystem disturbances

The challenge of anomaly detection

But even in these common use cases, above, there are some drawbacks to anomaly detection. From a conference paper by Bram Steenwinckel:

“Anomaly detection (AD) systems are either manually built by experts setting thresholds on data or constructed automatically by learning from the available data through machine learning (ML).”

It is tedious to build an anomaly detection system by hand. This requires domain knowledge and—even more difficult to access—foresight.

For an ecosystem where the data changes over time, like fraud, this cannot be a good solution. Building a wall to keep out people works until they find a way to go over, under, or around it. When the system fails, builders need to go back in, and manually add further security methods.

Under the lens of chaos engineering, manually building anomaly detection is bad because it creates a system that cannot adapt (or is costly and untimely to adapt).

Anomaly detection with ML

Machine learning, then, suits the engineer’s purpose to create an AD system that:
Works better
Is adaptive and on time
Handles large datasets
Despite these benefits, anomaly detection with machine learning can only work under certain conditions.

Anomaly detection in three settings

In a 2018 lecture, Dr. Thomas Dietterich and his team at Oregon State University explain how anomaly detection will occur under three different settings. They all depend on the condition of the data.

The three settings are:
Supervised: raining data is labeled with “nominal” or “anomaly”.
Clean: In the Clean setting, all data are assumed to be “nominal”, and it is contaminated with “anomaly” points.
Unsupervised: In Unsupervised settings, the training data is unlabeled and consists of “nominal” and “anomaly” points.

Pen Testing Systems

AI Pen Testing Systems

The term “AI” is often used carelessly when describing software tools. Here we identify a small number of the testing systems available with some level of genuine AI included. Deep Exploit learns how to exploit a system by carrying out attacks using learned methods and by brute force. It carries out a wide-range attack, targeting all open ports, using traditional attack methods. It can then focus its approach by targeting a specific port number and application, using its arsenal of exploits and payloads. It learns by assessing feedback from successful attacks.

Pentoma assesses servers and applications to find security risks such as:

SQL Injection;

File inclusion;

Unvalidated Forwards and Redirects;

Cross-site scripting (XSS).

It uses ML and AI (to an extent) to evolve and grow its assessments and techniques.

Wallarm uses nodes deployed in the cloud network to provide dynamic protection against the most common application vulnerabilities (known as the OWASP top 10) including injection, broken authentication, sensitive data exposure and XML external entities. It can discover network assets, scan for vulnerabilities and monitor abnormal patterns. It learns application vulnerabilities using automated threat verification. Having blocked a malicious request, it mimics it to test the behaviour of the application.

- **Hackers Using AI**
- The use of AI is not confined to application development and operations; hackers are using AI to assist their activities. Indeed, AI can itself be hacked. The algorithm at the heart of the AI process can be manipulated during its learning phase and after deployment. Security specialist Darktrace reports that AI-driven malware is being used to mimic the behaviour of a human attacker, increasing the stealth and scalability of attacks. By extending malware such as TrickBot, hackers can adopt contextual awareness. An AI-based attack can autonomously assess the target and determine how to avoid detection. This makes it much harder to track the criminal behind it.
-
- Human Pen Testers v AI
- It’s unlikely that we will adopt the Luddites’ methodology of attacking our machines. While ML can learn from data, it’s not a substitute for a human Pen tester. AI has the potential to disrupt the industry further, but it still has some way to go. Realistically, by taking on the routine tasks, ML should make our careers as Pen testers more enjoyable, giving us time to communicate and to think outside the box.

SOCIAL ENGINEERING

Social engineering is the art of manipulating people so they give up confidential information. The types of information these criminals are seeking can vary, but when individuals are targeted the criminals are usually trying to trick you into giving them your passwords or bank information, or access your computer to secretly install malicious software–that will give them access to your passwords and bank information as well as giving them control over your computer.

Criminals use social engineering tactics because it is usually easier to exploit your natural inclination to trust than it is to discover ways to hack your software. For example, it is much easier to fool someone into giving you their password than it is for you to try hacking their password (unless the password is really weak).

Security is all about knowing who and what to trust. It is important to know when and when not to take a person at their word and when the person you are communicating with is who they say they are. The same is true of online interactions and website usage: when do you trust that the website you are using is legitimate or is safe to provide your information?

Ask any security professional and they will tell you that the weakest link in the security chain is the human who accepts a person or scenario at face value. It doesn’t matter how many locks and deadbolts are on your doors and windows, or if have guard dogs, alarm systems, floodlights, fences with barbed wire, and armed security personnel; if you trust the person at the gate who says he is the pizza delivery guy and you let him in without first checking to see if he is legitimate you are completely exposed to whatever risk he represents.

What Does a Social Engineering Attack Look Like?

Email from a friend

If a criminal manages to hack or socially engineer one person's email password they have access to that person's contact list–and because most people use one password everywhere, they probably have access to that person's social networking contacts as well.

Once the criminal has that email account under their control, they send emails to all the person's contacts or leave messages on all their friend's social pages, and possibly on the pages of the person's friend's friends.

Taking advantage of your trust and curiosity, these messages will:

Contain a link that you just have to check out–and because the link comes from a friend and you're curious, you'll trust the link and click–and be infected with malware so the criminal can take over your machine and collect your contacts info and deceive them just like you were deceived

Contain a download of pictures, music, movie, document, etc., that has malicious software embedded. If you download–which you are likely to do since you think it is from your friend–you become infected. Now, the criminal has access to your machine, email account, social network accounts and contacts, and the attack spreads to everyone you know. And on, and on.

Email from another trusted source

Phishing attacks are a subset of social engineering strategy that imitate a trusted source and concoct a seemingly logical scenario for handing over login credentials or other sensitive personal data. According to Webroot data, financial institutions represent the vast majority of impersonated companies and, according to Verizon's annual Data Breach Investigations Report, social engineering attacks including phishing and pretexting (see below) are responsible for 93% of successful data breaches.

Using a compelling story or pretext, these messages may:

Urgently ask for your help. Your ’friend’ is stuck in country X, has been robbed, beaten, and is in the hospital. They need you to send money so they can get home and they tell you how to send the money to the criminal.

Use phishing attempts with a legitimate-seeming background. Typically, a phisher sends an e-mail, IM, comment, or text message that appears to come from a legitimate, popular company, bank, school, or institution.

Ask you to donate to their charitable fundraiser, or some other cause. Likely with instructions on how to send the money to the criminal. Preying on kindness and generosity, these phishers ask for aid or support for whatever disaster, political campaign, or charity is momentarily top-of-mind.

Present a problem that requires you to "verify" your information by clicking on the displayed link and providing information in their form. The link location may look very legitimate with all the right logos, and content (in fact, the criminals may have copied the exact format and content of the legitimate site). Because everything looks legitimate, you trust the email and the phony site and provide whatever information the crook is asking for. These types of phishing scams often include a warning of what will happen if you fail to act soon because criminals know that if they can get you to act before you think, you’re more likely to fall for their phishing attempt.

Notify you that you’re a ’winner.’ Maybe the email claims to be from a lottery, or a dead relative, or the millionth person to click on their site, etc. In order to give you your ’winnings’ you have to provide information about your bank routing so they know how to send it to you or give your address and phone number so they can send the prize, and you may also be asked to prove who you are often including your social security number. These are the ’greed phishes’ where even if the story pretext is thin, people want what is offered and fall for it by giving away their information, then having their bank account emptied, and identity stolen.

Pose as a boss or coworker. It may ask for an update on an important, proprietary project your company is currently working on, for payment information pertaining to a company credit card, or some other inquiry masquerading as day-to-day business.

Tips to Remember:

Slow down. Spammers want you to act first and think later. If the message conveys a sense of urgency or uses high-pressure sales tactics be skeptical; never let their urgency influence your careful review.

Research the facts. Be suspicious of any unsolicited messages. If the email looks like it is from a company you use, do your own research. Use a search engine to go to the real company’s site, or a phone directory to find their phone number.

Don’t let a link be in control of where you land. Stay in control by finding the website yourself using a search engine to be sure you land where you intend to land. Hovering over links in email will show the actual URL at the bottom, but a good fake can still steer you wrong.

Email hijacking is rampant. Hackers, spammers, and social engineers taking over control of people’s email accounts (and other communication accounts) has become rampant. Once they control an email account, they prey on the trust of the person’s contacts. Even when the sender appears to be someone you know, if you aren’t expecting an email with a link or attachment check with your friend before opening links or downloading.

Beware of any download. If you don’t know the sender personally AND expect a file from them, downloading anything is a mistake.

Foreign offers are fake. If you receive an email from a foreign lottery or sweepstakes, money from an unknown relative, or requests to transfer funds from a foreign country for a share of the money it is guaranteed to be a scam.

Ways to Protect Yourself:

Delete any request for financial information or passwords. If you get asked to reply to a message with personal information, it's a scam.

Reject requests for help or offers of help. Legitimate companies and organizations do not contact you to provide help. If you did not specifically request assistance from the sender, consider any offer to 'help' restore credit scores, refinance a home, answer your question, etc., a scam. Similarly, if you receive a request for help from a charity or organization that you do not have a relationship with, delete it. To give, seek out reputable charitable organizations on your own to avoid falling for a scam.

Set your spam filters to high. Every email program has spam filters. To find yours, look at your settings options, and set these to high–just remember to check your spam folder periodically to see if legitimate email has been accidentally trapped there. You can also search for a step-by-step guide to setting your spam filters by searching on the name of your email provider plus the phrase 'spam filters'.

Secure your computing devices. Install anti-virus software, firewalls, email filters and keep these up-to-date. Set your operating system to automatically update, and if your smartphone doesn't automatically update, manually update it whenever you receive a notice to do so. Use an anti-phishing tool offered by your web browser or third party to alert you to risks.

Webroot's threat database has more than 600 million domains and 27 billion URLs categorized to protect users against web-based threats. The threat intelligence backing all of our products helps you use the web securely, and our mobile security solutions offer secure web browsing to prevent successful phishing attacks.