Malware Defense Through Collaborative Filtering

White Paper

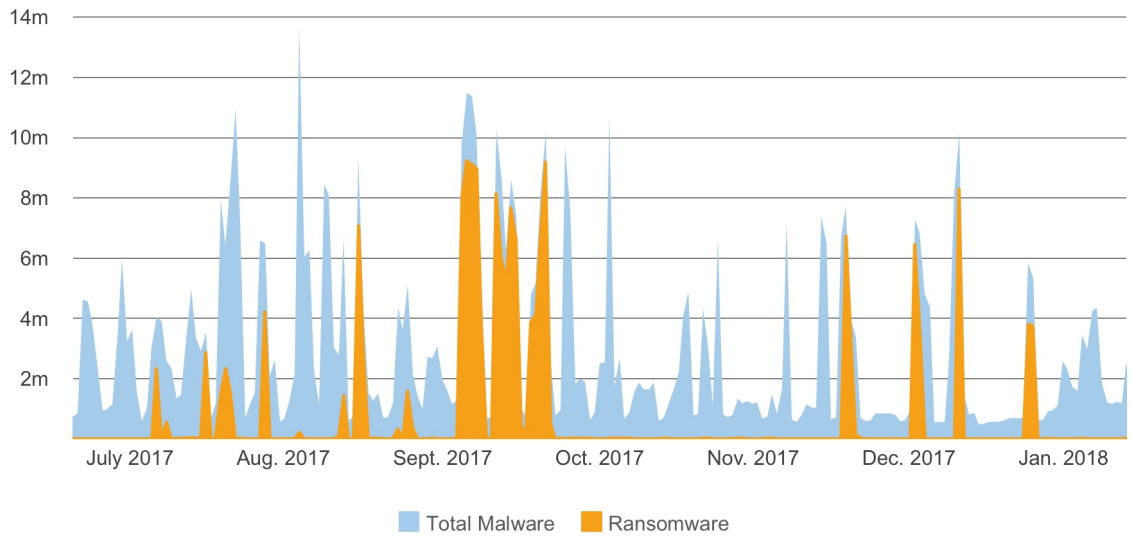
April 20, 2018

Abstract

Traditional anti-virus solutions rely on heuristics and signature-based detection, which derives from the outdated assumption that the malware you saw yesterday will look the same way today. Malware that cyber criminals create has evolved significantly in the last few years. Modern malware authors change their tactics constantly and adapt accordingly to evade detection. The majority of advanced malware samples seen in the wild today are in active campaigns for approximately an hour while some unique samples are beginning to be seen only once. This short lifespan makes it difficult for traditional anti-virus solutions to catch active threats and they struggle to keep pace. Sandboxing solutions that "detonate" and observe samples would provide significantly better effectiveness, but these solutions require an abundance hardware of resources in even enterprise environments and therefore are not cost effective to run in much larger service provider and hosting provider environments.

One attack type that has increased dramatically in frequency is "Ransomware", a malware type that encrypts a user's data, holding it for ransom, with the hope of extorting payment from the user to decrypt the data. Ransomware exploded in popularity in 2016 with 4,000 ransomware attacks per day, according to IBM.[1] The Locky family of ransomware took position as a dominant threat and until recently, was one of the biggest threats the day. Other notable Ransomware families that have recently been in the wild include GandCrab, WannaCry, and Petya/Notpetya.

2017-2018 Volume of Malicious vs Ransomware Traffic as designated by Cloudmark Authority.

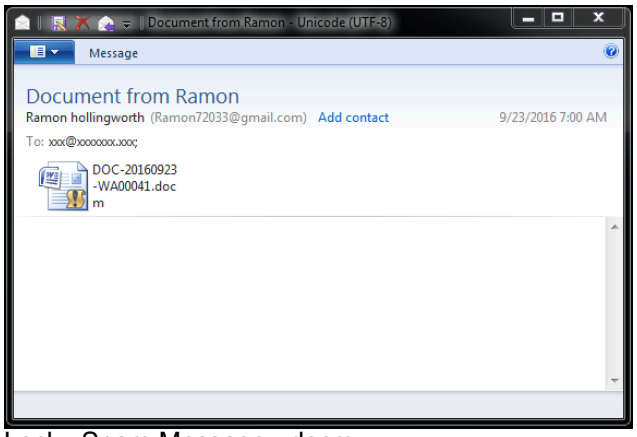


[1] Ransomware: How Consumers and Businesses Value Their Data. IBM X-Force Research, 2016.

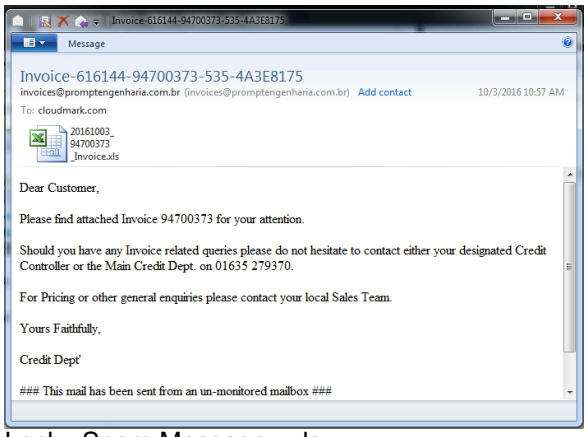
Ransomware: Locky

There are many Ransomware versions, but until recently, Locky was by far the most dominant player in the Ransomware landscape. Its delivery was characterized by high volume spam campaigns via email. These malicious campaigns, disguised as a trustworthy source, trick users into opening and executing an infected script-based file (.js) or macro-enabled document (.docm, .dotm, .xls, etc.). Once a computer is infected, it encrypts many different file types on a machine, such as, Word docs, Excel, PowerPoint, images, as well as Bitcoin wallets, database files, source code, etc. The ransom or fee is usually paid in some form of cryptocurrency. Cybersecurity Ventures predict global ransomware damage will increase to $1.5 billion in 2019, up from $325 million in 2015. The rising costs are a result of an increase in frequency and sophistication of attacks – 1 attack every 14 seconds by 2019[2].

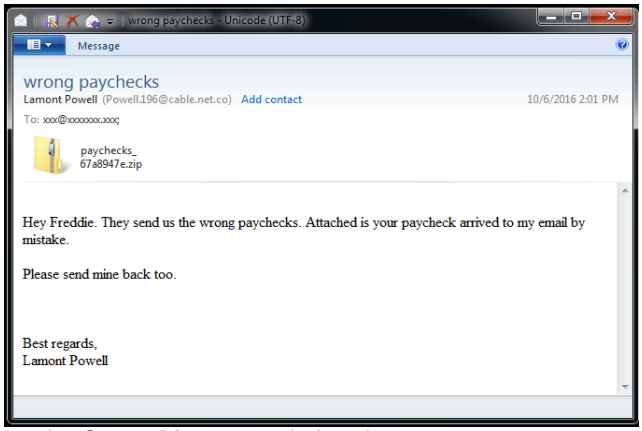
Types of Ransomware Messages: Locky



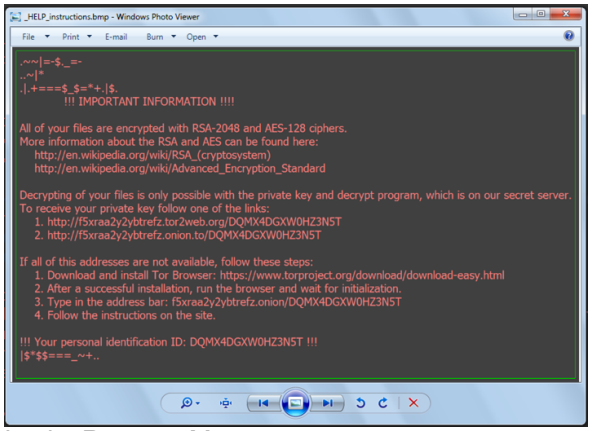
Locky Spam Message: .docm



Locky Spam Message: .xls



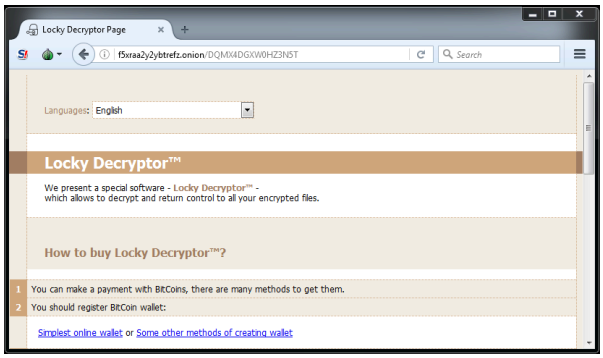
Locky Spam Message: .js in .zip

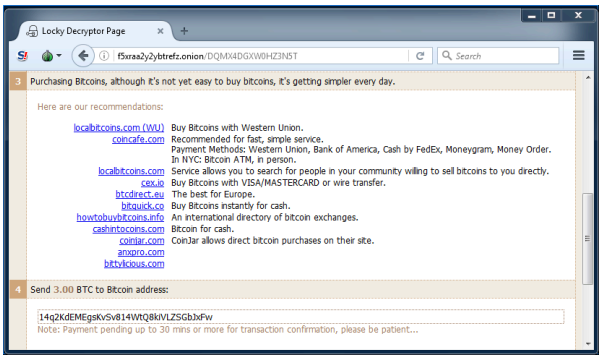


Locky Ransom Message

[2] Ransomware Damage Report, Cybersecurity Ventures.

Locky Payment Page





Handling Malware with Cloudmark Authority and The Global Threat Network

Through the power Cloudmark Authority and securing 400+ million messaging and email inboxes worldwide, Cloudmark provides a unique malware identification and detection solution far beyond the capabilities of traditional anti-virus services. Our algorithms allow our engine to combine detailed information about the content and structure of attachments with key data about the structure and content of the email message. This enables Cloudmark Authority to generate meaningful fingerprints, without having to do a deep analysis on the attachment during the initial scan. Deep analysis is included through the Global Threat Network feedback system, an agentless approach to dynamic malware analysis utilizing a reputation-based, trusted community of real-time honeypots and end users to manage and evaluate malicious traffic, then translate into high performance fingerprints. Fingerprint updates are pushed out to Cloudmark Authority every 15 seconds and more comprehensive changes to the message is pushed out globally within minutes.

Cloudmark's fingerprinting algorithm circumvents the time-intensive "reverse engineering" analysis of conventional technologies allowing its system to identify and squelch new malware strains in near real-time. The Cloudmark technology is language-agnostic, format-agnostic, representation-agnostic, and protocol-agnostic — making it particularly suited to combat all forms of malicious content.

Individuals who use messaging services, such as email, SMS, and MMS, are valuable resources in discerning the difference between content that is spam and content that is legitimate. When users are pooled together and allowed to "vote" on which content is and is not spam, the quality of their opinions is astonishingly high. The concept of pooling individual opinions regarding a piece of data is known as collaborative filtering.

The actual process of collaborative filtering is complex, but intuitive, and can be broken down into several components: fingerprint generation, weighted voting, fingerprint acceptance, and filtration.

Fingerprint Generation

The fingerprint-generation phase begins with the computation of a mathematical function on each email message that is received by the client. The output of the function, known as the fingerprint, is a numerical value that can be created only from that exact email and specifically associated variations of that email. By generating fingerprints that are flexible to small variations, Cloudmark makes it extremely difficult for spammers to change a slight amount of text in order to evade previously generated fingerprints.

Weighted Voting

Fingerprints are submitted to a central database during the weighted voting period. A reporter (e.g. an end user, system administrator, spamtrap, or another automated system) identifies suspicious behavior and the reporter votes on whether a message is spam by submitting the fingerprint with a flag indicating the possibility of that content can be spam. The weight assigned to the reporter's vote is based upon the reporter's historical trustworthiness in correctly reporting spam. Therefore, if a reporter has a hard time determining if a piece of content is spam or not spam, the vote will ultimately be weighted with less importance. Once the number of weighted votes on a specific fingerprint reaches a predetermined threshold, the fingerprint is accepted as a verified indicator of spam. From that point forward, any reporter who checks content which contains that fingerprint, the database will be informed that the community has determined the content to be spam, and the system will filter out the message.

Fingerprint Acceptance

The algorithm for weighting user votes, which determines the correct point at which to accept a fingerprint, and which handles disagreement inside the community regarding the "spamminess" of a message, is encoded in the Cloudmark Trust Evaluation System, or TES. Just as in a real-world human community, individuals who behave well by quickly and correctly identifying spam become trusted and are rewarded by having their opinions weighted more heavily in the future.

Filtration

The principle of community-based collaborative filtering was critical in the design of our anti-spam service, known as the Cloudmark Global Threat Network [1]. The Cloudmark service is not limited to the collaborative filtering of email. From the outset, the Cloudmark service has been designed to implement a general framework for filtration, based on collaborative opinion. The Cloudmark approach can be rather easily adapted to almost any messaging medium by enabling feedback and filter hooks into the medium.

Anti-Virus

At its core, traditional anti-virus (AV) software provides signature-based detection of malware's unique segment of code and then cross-checks the signature with a database of known viruses. Virus signatures are collected by several methods, including user submissions, and examining honeypots, or by machines solely dedicated to accepting and recording malware attacks for later analysis. The latency associated with the collection and analysis phase can be several hours.

Cloudmark addresses the problem of rapid response and precise identification of viruses by leveraging its massive installed base of human malware reporters (over 400 million inboxes) and by introducing a special fingerprinting algorithm for capturing executable code. Since any one of Cloudmark's many users may potentially be the first person to see an emergent malware message, our installed base effectively becomes the world's largest network of honeypots for malicious email content, dramatically reducing our virus sample collection time.

This fingerprinting algorithm automates the "reverse engineering" process used by conventional anti-virus vendors, and successfully removes many of the obfuscations hackers place on binaries to evade detection. By automating this process, Cloudmark's technology essentially removes the arduous, human-driven task of malware analysis. In practice, the Cloudmark anti-virus engine is able to detect and filter viruses as soon as users click "block," which is well before they are even named by the security community.

Cloudmark scans up to **6 billion email messages** per day at the content filter level. When malware creators launch a new attack, the combined power of all the Cloudmark fingerprinting algorithms, provides significant coverage to detect the malware from multiple angles and allows Cloudmark Authority to quickly respond and stop new attacks.

Threats of the Future

Advanced societies have evolved a standard methodology for dealing with new threats. They come together and determine, through consensus, the nature of, and appropriate response to, the threat. The Cloudmark service is a generic codification of this process, where users can submit fingerprints for any form of content, including malware and newer forms of ransomware, then vote upon the disposition of the content. Due to its content agnostic engineering, our technology can be extended to combat all forms of malicious content, ranging from today's threats to those which have not yet been designed.

References:

1. V.V. Prakash and A.O'Donnell. Fighting spam with reputation systems. Queue, 3 (9):36-41, 2005.