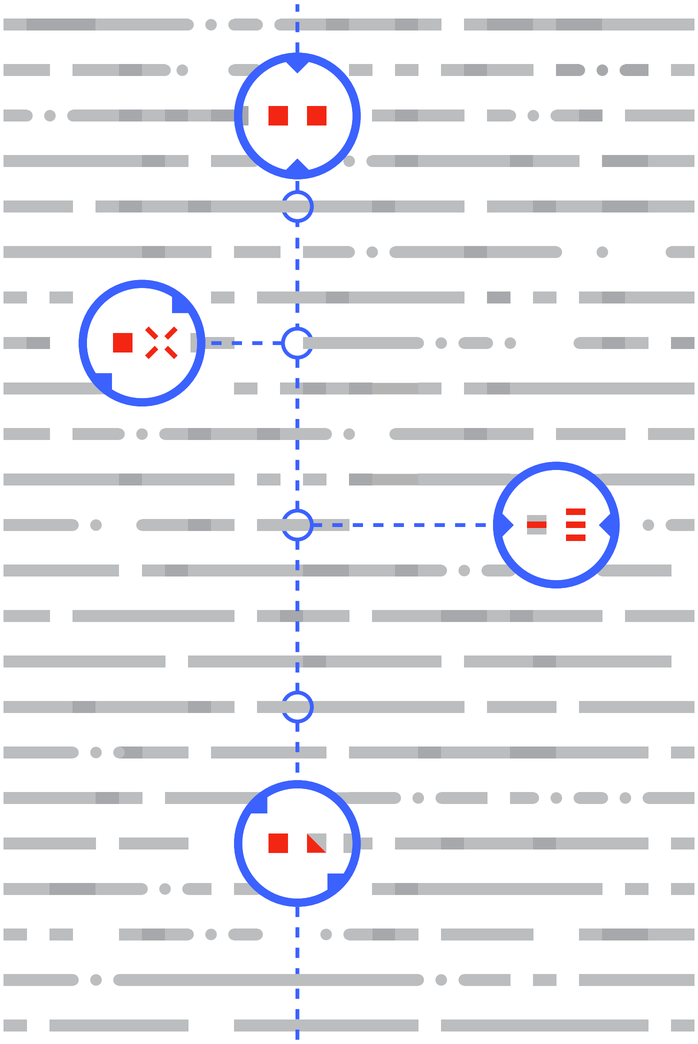
Outsmart Malware

VirusTotal Intelligence lets you hunt for malware threats that affect your networks.

VirusTotal Intelligence is one of the world's largest malware intelligence services. Security professionals rely on it to better understand the effects of malware in enterprise networks..



VirusTotal is one of the world's largest malware intelligence services. Millions of people use it every day to perform basic research on malware. This typically takes the form of checking aggregated verdicts from the AV products that participate in the VirusTotal service. While the VirusTotal name is well-known, it's likely that most users only know VirusTotal for its basic, verdict-focused public service. However, the VirusTotal platform also offers advanced and powerful malware intelligence services designed to help analysts perform better understand threats both in the wild and in their networks.

This advanced service is VirusTotal Intelligence (VTI), and it includes a variety of useful threat intelligence capabilities. This brief discusses three of these in more detail.

VIRUSTOTAL INTELLIGENCE

VTI is an intelligence layer across the billions of files within the VirusTotal database. VTI is sometimes described as a combination of Google and Facebook, applied to malware. VTI is "Google-like" in that it enables threat analysts to perform fast searches against a massive dataset, using multiple parameters to find malware samples that match a specific set of criteria. Criteria can includes malware families, labels, binary properties, metadata, etc.

VTI is also a bit like Facebook because for every malware match, it returns a detailed profile of that malware file. The profile includes key information such as: relationships (this malware sample to others), behavior, signals and metadata extracted from various tools, and detection data -- as well as crowdsourced information via analyst comments. Together, VTI acts as both a telescope ("show me broad malicious behaviors on the internet") and as a microscope ("...now show me the details of this particular malware file.")

VTI does this by executing a myriad of backend processes, including:

- Running tools in the background to extract suspicious signals from a sample. For example, PDFid, which indicates whether a pdf file contains javascript or invalid cross referenced sections, i.e. it will flag this pdf as suspicious.
- Correlating the samples coming into VirusTotal with other samples in the collection. It decompress a sample, examine the inner files and make parent-child connections.
- Scanning URLs to identify that a file has been seen at this location in the wild. If you upload an email attachment, VTI can tell you that this file is in the VirusTotal dataset, and here is where else we see this file.
- Inspecting recorded network traffic for every PCAP recorded network trace, to extract files seen on the wire and to identify network traces where a particular file has been seen elsewhere.
- Executing the samples in sandboxes to see what the malware actually does in a user system. VTI traces the actions that the sample performs and the communications it executes with network points to collect stolen data, etc.
- And more...

USEFUL VIRUSTOTAL INTELLIGENCE FUNCTIONS

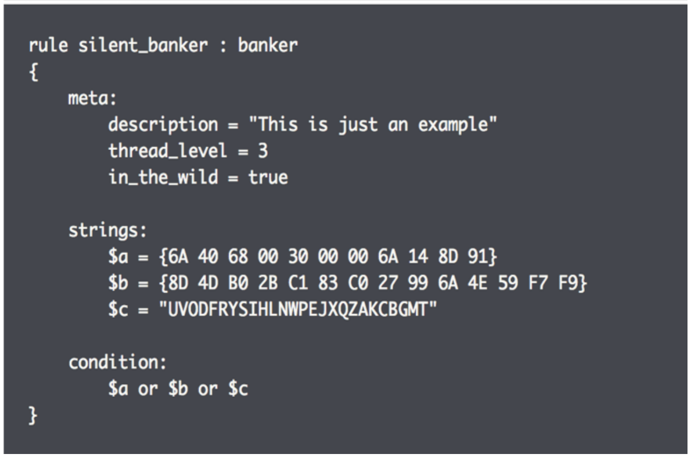
VTI has multiple threat intelligence functions, each of which helps enterprise security teams better understand the effects of malware in their networks. Three of the most interesting functions are malware threat hunting, clustering analysis, and relationship/behavior visualization.

#1: Malware Threat Hunting

Threat response and research teams need to understand the types of malware that are likely to affect their organizations. Examples might include banking trojans targeted at retail banks, ransomware targeting hospitals, etc.

VTI's threat hunting function helps security teams find malware samples and understand the signs of those samples' operations. With this information, threat response teams can update their infrastructure and better protect their networks. A good analogy to VTI threat hunting is Google Alerts, which allows someone to enter a given string into the system and get notified as soon as an article with a matching string is indexed.

VTI has a roughly similar capability, YARA, which is a pattern matching language for building both simple and complex rules to identify variants of malware pertaining to a family or attacker groups. Users can create their own rules, enter them into VTI, and those rules will find new samples coming into VirusTotal and alert the users. It's effectively "Google Alerts for malware," and enables analysts to see how attackers are evolving their techniques and to keep tabs on particular attackers and campaigns.

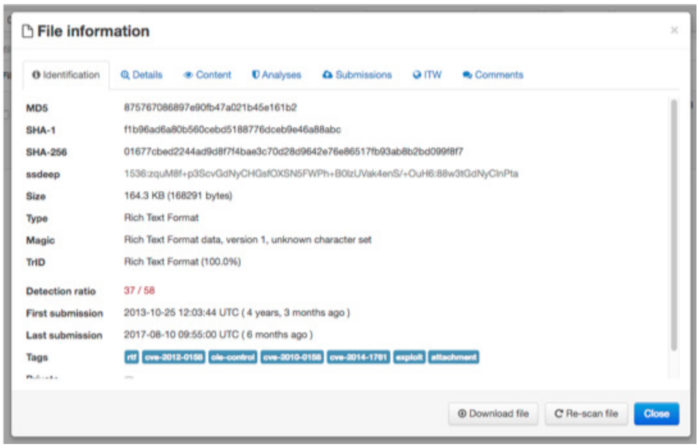


```
rule silent_banker : banker
{
    meta:
        description = "This is just an example"
        thread_level = 3
        in_the_wild = true

    strings:
        $a = {6A 40 68 00 30 00 00 6A 14 8D 91}
        $b = {8D 4D B0 2B C1 83 C0 27 99 6A 4E 59 F7 F9}
        $c = "UVODFRYSIHLNWPEJXQZAKCBGMT"

    condition:
        $a or $b or $c
}
```

When an analyst finds a specific family or instance of targeted malware, a natural next step is to look back in time to see the sample's history and how it has evolved. Old samples might leave artifacts behind or might have relationships with other files in the database, and it's useful to examine them. For that purpose, VTI includes Retrohunt, which provides YARA capability applied back in time against the VirusTotal file dataset. This is useful because certain types of malware might go unnoticed for years. Today, a new detection technique might find the malware, and an analyst might want to see if any variants existed in the past but were undetected.



MD5	875767086897e90fb47a021b45e161b2
SHA-1	f1b96ad6a80b560cebd5188776dceb9e46a88abc
SHA-256	01677cbed2244ad9d8f7f4bae3c70d28d9642e76e86517fb93ab8b2bd099f8f7
ssdeep	1536:zquM8f+p3ScvGdNyCHGsfOXSN5FWPh+B0lzUVak4enS/+OuH6:88w3tGdNyClnPta
Size	164.3 KB (168291 bytes)
Type	Rich Text Format
Magic	Rich Text Format data, version 1, unknown character set
TrID	Rich Text Format (100.0%)
Detection ratio	37 / 58
First submission	2013-10-25 12:03:44 UTC (4 years, 3 months ago)
Last submission	2017-08-10 09:55:00 UTC (6 months ago)
Tags	rtf cve-2012-0158 ole-control cve-2010-0158 cve-2014-1761 exploit attachment

For example, Stuxnet went unnoticed for years. Once it was eventually discovered, researchers examined the VirusTotal library and found variants of Stuxnet that had been archived for years, unnoticed. So, as techniques change, and detection techniques change, it's useful to look at the historical record and see if a recently-detected sample was in the wild earlier, and if so, what did it look like.

#2: Cluster Analysis

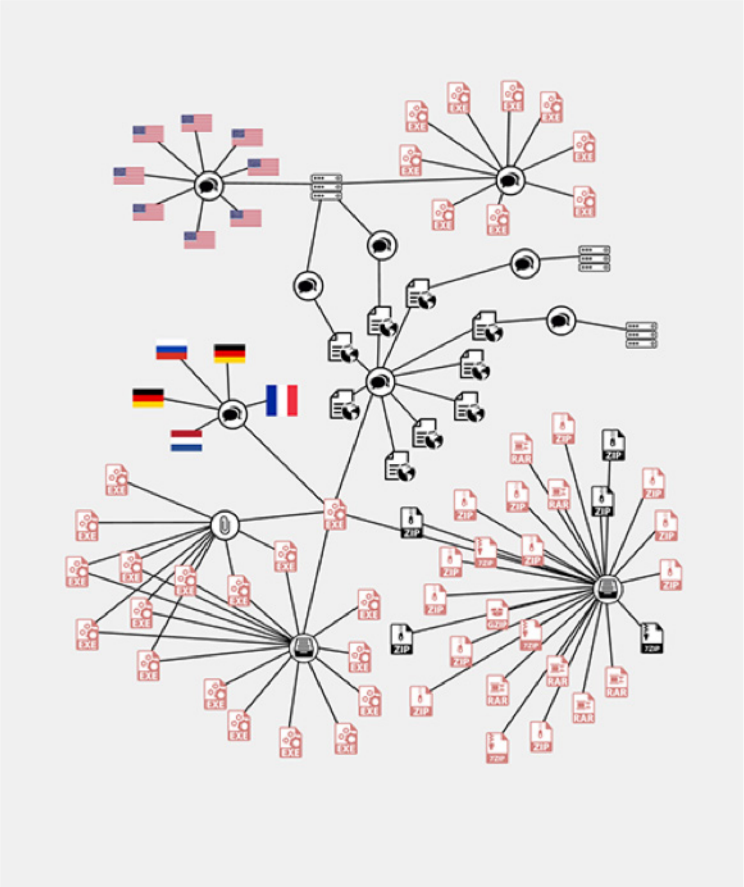
VTI provides an internally-developed function that classifies malware samples into groups based on "family resemblance." Why would malware be in a family and why is this important? Cybercriminals often use kits to create malware -- these make it easy to create a new variant of malware, cheaply. Instead of creating each new variant from scratch, a kit allows a criminal to enter a few parameters (e.g. I want to exfiltrate data to this URL, I want to grab documents from the endpoint, etc.), and then the kit generates a new piece of malware, slightly different from others and tuned to a criminal's particular campaign goals.

The result of wide use of malware generation kits is that many different malware binary files share essentially the same behavior, since they are from the same family. This is important because more than one million sample files are uploaded to VirusTotal per day, but 40-60% of those come from same ~200 clusters (i.e. families). While one million files per day will likely overwhelm a team's analysis capabilities, examining 200 clusters is more manageable. To create clusters, VTI analyzes the malware characteristics that tend not to change across individual variants. For example, most Windows malware might share certain techniques, PDF malware might share certain characteristics, etc. VTI regularly updates and evolves its algorithms to support effective clustering.

For enterprise threat analysts, it is easier to build rules for the 200 major malware clusters than for one million files. By building cluster-level rules, analysts gain coverage for up to 60% of all incoming files. As a secondary step, researchers can then write rules for the small clusters that might affect only their own organization. This allows a threat research team to focus on the malware that is going under the radar within that team's organization. An example might be spearfishing, targeting only a specific company's executives. VTI clustering enables threat teams to look at small clusters and uncover interesting files and activity within them.

#3: Relationship and Behavioral Visualization, i.e. VTI Graph

A recently-added VTI feature is relationship and behavioral visualization, called VTI Graph. VTI Graph allows you to visualize all the relationships within VTI, and quickly conveys what's going on with a piece of malware. Without VTI Graph, an analyst must repeatedly open up every profile of a malware sample, keep tabs open for each of them, cut and paste information into a master document, and try to understand the commonalities of the threats studied. In contrast, with VTI Graph it is easy to see that many different samples are pointing to the same domain, for example, a command and control domain. This is hard to see when looking at each one in a list of matches, as each sample may communicate with multiple domains, making it difficult to spot the common domains. VTI Graph stitches it all together -- the analyst sees instantly that 200 samples are communicating with a specific domain. The security team can then update its infrastructure rules to block communications to that domain.



Without VTI Graph, users are required to build the map on their own, copy paste into a doc or write a script to figure out how the files operate and communicate. With VTI Graph, an analyst gets the whole map at once. While making maps might be fun while playing adventure games, it's much less enjoyable during threat hunting. VTI Graph provides two benefits: productivity and accuracy. Analyst productivity is much higher because tedious manual steps are eliminated, and accuracy is higher, since a manual analysis may miss steps and information. With VTI Graph, malware behavior analysis is instant and more accurate via automation.

In addition, an analyst can then store the graph, i.e. the map of the investigation, within VTI. VTI Graph saves it and provides a link which can be shared with other users. Previously, the analyst would have to give peers a list of steps to be manually performed in VTI, so that they could recreate the relationship-behavior map. Instead, peers get the information instantly. The graph is actually a visualization of a data structure that VTI builds under the hood, so VTI can also export this structure to another tool for additional analysis. Finally, customers can layer access control over this structure, to ensure data privacy.

VTI continues to extend Graph's functionality. Current research and development efforts include integration of VTI Graph with the clustering analysis described above. Consider a case where an analyst is examining a graph and sees a node that has little context information. It would be quite useful to display all similar (i.e. clustered) files and to check if one of those files has enough context to shed light on the original node file. In addition, the VTI team is also adding common node types, such as scripts, attacker, victim, email, campaigns, (i.e. nodes that give the entire picture) beyond files and URLs.

A VTI USE CASE: ANALYZING BANKING TROJANS

Let's use an example use case to illustrate some of the features above: Banking trojans targeting the financial services industry. Banking trojans are designed both to steal user credentials and to execute hidden transactions to steal money from users during normal banking flow. Since a trojan executes within a user's browser, it can alter balances and hide transactions from users. This is bad for banking customers, since they lose money, and bad for banks because of hits to reputation and the legal requirement to repay the losses.

In this example, a bank's incident response team uses VTI to hunt for samples of trojans created for banking theft. VTI allows the analyst to search with labels, i.e. show me all samples detected as password stealers, those that modify the system's hosts file, those that contain strings relative to banks, etc. The team may also follow security companies' blogs and get notified that certain samples have been uploaded to VT. The bank's analyst team can download one specimen from each family, and build an encyclopedia of applicable samples. The analysts can use multiple samples to see common patterns, then create YARA rules to identify these.

The analysts load those rules into VTI malware hunting, and be notified for each new instance of any malware in those families. This allows the team to automate the flow, and for each notification, to download a sample and apply a set of plug-ins to decipher their config files and reveal which network infrastructure the credentials are sent over, and the domains from which additional payload files are downloaded. The bank updates its own security controls to block these files and network points.

The bank can also notify ISPs and other providers to shut down the malicious infrastructure. This renders many types of malware ineffective. Because of server side polymorphism, there may be thousands of variations of any type malware, all using just a few network points. By identifying and shutting those down, it is possible to stop entire families of malware aimed at attacking banking customers. The bank, by shutting down just a single piece of network infrastructure, can prevent fraud in thousands of infected machines.

A similar action can be taken for spearfishing attacks. A firm's threat response team can access relevant samples by searching in VTI, then build rules in malware hunting, then track how the malware evolves. This enables the analysts to perform attribution and to understand how attackers are operating. For example, the team might analyze which email subjects and bodies are common across attacks, then put rules on its own email servers to quarantine these messages.

All of the above is available through both API and through a GUI. In this way, VTI allows users to test rules and techniques easily, and then to automate via scripting with the API.

CONCLUSION

This brief describes several advanced malware intelligence capabilities, delivered through VTI advanced searching, malware hunting, clustering analytics, Retrohunt, and relationship visualization. These capabilities give threat response teams a rapid way to understand malware attacks that might affect their companies, as well as the information needed to update their own security infrastructure to make use of this data.

In practice, many enterprise security teams aren't familiar with these capabilities. They are familiar with VirusTotal's basic research features, and may upload malware samples via the free virustotal.com site. However, since malware continues to be among the most pervasive and damaging threats to cybersecurity, enterprise threat analysts find VirusTotal Intelligence's advanced analytics tools to be useful in understanding and responding to malware attacks. VirusTotal continues its dual mission of providing ever-increasing value to security partners while providing advanced intelligence tools to end customers.

If you are interested in learning more about or testing the capabilities described above, please complete the Contact Us form on the virustotal.com website, or send an email to info@virustotal.com.