WHITE PAPER

McAfee
Together is power.

Integrating SIEM into Your Threat Hunting Strategy

Table of Contents

3 What Are Data Sets in the Context of Threat Hunting?
4 Building a Threat Hunting Toolkit
5 Getting an Overall Picture with Maltrail
6 Adding Layers of Detail with Sinkholes
7 A Typical Threat Hunt
10 Summary
11 Strategy, tactics, and operations

Integrating SIEM into Your Threat Hunting Strategy

Cyberthreat hunting is the process of proactively and iteratively searching through networks and data sets to detect threats that evade existing automated tools.[1] While that sounds straightforward, it is fraught with complexities that are neither obvious nor easy to remedy. For example, what are the data sets? Where do they come from? How do you search through them iteratively? How can you be proactive?

In this paper, we offer both an approach and a toolkit for threat hunting. We show you how to aggregate and correlate the data your tools provide into a single analysis tool—an advanced security information and event management (SIEM) platform—to detect and block cyberthreats. We show you how a solid threat-hunting infrastructure can help you achieve the proactive goal of the definition and how to advance the proactive defense infrastructure of your enterprise. While the centerpiece for your threat-hunting toolkit is an SIEM, we will use some open source tools to collect data and show how commercial tools can fit in as well.

Remember, threat hunting is a team sport. Sharing results of your hunts with other hunters—perhaps using different tools—can only gather more information for you both. Also, and equally important, there is a lot of data, and that means that you could take a lot of time to sift through it and get useful results. Anything that you can do to shorten the hunting cycle without sacrificing accuracy or thoroughness is a good thing.

What Are Data Sets in the Context of Threat Hunting?

Experienced threat hunters have their data set preferences, but what is most important is defining the types of data you are seeking. The overall objective of your threat management strategy will dictate, to a large degree, what types of data you need. The data dictate the data sets, and the data sets dictate the tools.

There is a misconception that you should start with the tools and work the other way. However, if you don't know what you're looking for, how can you know what tool to use to find it? Additionally, do you want to be able to apply forensic analysis to your data? The answer to that is usually "yes," but that affirmative opens up a new level of complexity.

There is a misconception that you should start with the tools and work the other way. However, if you don't know what you're looking for, how can you know what tool to use to find it?

Generally, we think of the following types of data sets as useful for threat hunters:

- Resources on hosts and endpoints such as PowerShell transcripts, logs, and more
- Firewall and intrusion detection systems (IDS)/ intrusion prevention systems (IPS) logs
- Malware lists and captures
- Passive DNS
- Whois
- Web logs (access, proxy, referrer, and others)
- Process execution logs
- Authentication and Active Directory logs
- Registry modifications
- Syslog and Microsoft Windows event logs
- Netflow
- Network events
- Other security device logs
- Malicious domain lists
- Crowd-sourced malicious activity lists

There are many sources for these data, and you can access the data in a variety of ways. For example, there are simple ways to collect all malicious scans and attempts against your perimeter and compare that with the same type of data collected inside your enterprise. Some of those ways are free, so there is no need to extract that data from expensive tools such as IDS. That is not to say that IDS is not useful. What we are saying is this: select the right tool for the specific task.

Another important point is that more data is always better than less data. Never mind that huge data sets are tedious to analyze. Our tools will do that analysis for us. For example, a free tool called Maltrail will collect every attack/probe attempt against us. We set it on the outside perimeter of our test network. In a typical 24-hour period on our test network, with just one sensor exposed to the internet, it averages more than 6,000 events. Consider multiple sensors on a much larger footprint, such as we would see in a typical enterprise, and we likely would see well into the hundreds of thousands and, perhaps, millions, of events daily. The tool breaks that down for us and, feeding the output of the tool to an SIEM breaks it down even more, enabling us to do a cogent analysis. More important, Maltrail, on a typical day, might find one high-risk event and, perhaps, five or fewer medium risk events. The rest will be low.

Building a Threat Hunting Toolkit

To capture the data, you need a very comprehensive toolkit. That toolkit consists of cyberthreat intelligence feeds, in-house capture and logging systems, analysis tools, and correlation tools. In this section, we'll examine some of the available tools to stock your threat lab. As you become more integrated into threat hunting, you will develop additional favorites that you can add to the list, making it more personalized for you and your organization. You may, also, determine that some of the tools we discuss are not necessary for your environment.

Also, you should note that the tools that we are examining in this paper represent a samplIng of what is available. There are lots of different tools, and many of

them are quite effective. However, this toolkit will get you started and, most importantly, will show how you can integrate your threat hunting into a coherent toolkit and bring the data together in a single analysis tool: the SIEM.

Getting an Overall Picture with Maltrail

Maltrail is a free tool that detects malicious traffic by comparing data streams directed at its sensors with publicly available blacklists. It also looks at suspicious activity, anti-malware lists, and custom lists that you can define. It calls its finding "trails," and a trail can be an IP, domain name, a URL or URI or, really, just about any malicious indicator. At the time of this writing, there are more than 100 different trails that Maltrail watches.

Figure 1. Maltrail dashboard.

On our honeynet, with just four IPs exposed to the internet, we saw an event count on our SIEM of 1.2 million associated with Maltrail over a 24-hour period.

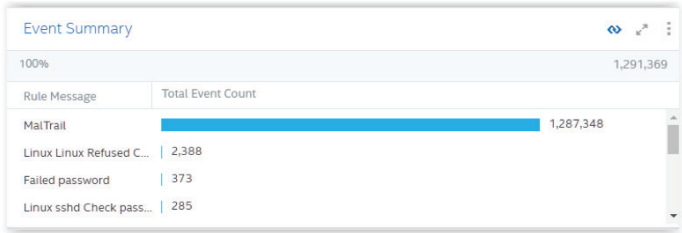

Figure 2. Maltrail events fed to McAfee Enterprise® Security Manager SIEM

Even without any other tools, this presents a prodigious amount of data to analyze. Fortunately, the SIEM provides some help. In Figure 3, we build a dedicated view of just the Maltrail traffic over the same 24-hour period.

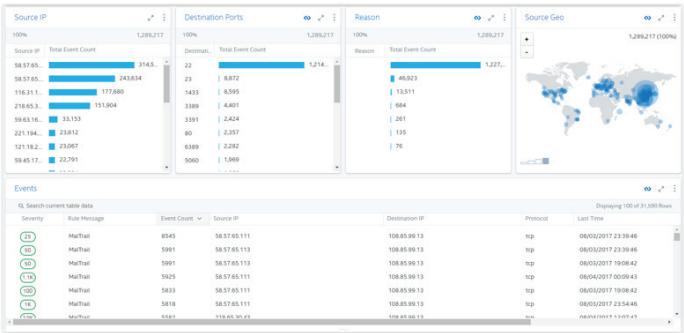

Figure 3. First level of analysis of Maltrail data.

There are several things here that might catch our eye. First, we see four IPs that seem to do the lion's share of attacks. Second, we see that virtually all of the attacks

are against port 22. Finally, the source geo is heavily centered in China. While this is interesting for context, we'd like a bit more to help us zero in on what we should be blocking or not. We also want to know if we have a threat active in our enterprise.

Adding Layers of Detail with Sinkholes

Sinkholes are spoofed DNS servers. They work by misguiding packets bound for a particular IP address or domain into a black hole, preventing it from communicating with the malicious endpoint, often a command and control (C&C) server. As an added benefit, you can collect information about what malicious actors are attempting to access your enterprise.

You populate your sinkhole with blacklists from a variety of sources. It is not uncommon to have tens of thousands of blacklisted addresses at any given time. When packets attempt to communicate with a destination on the blacklist, the sinkhole simply black holes it, and your enterprise is safe. A good example of this is when a user gets hooked by a phishing email and the malware attempts to communicate with its C&C. It gets black holed, and your enterprise is safe from that threat.

We use a freeware sinkhole created by Guy Bruno as a SANS project and we have found it very useful. However, since we use it for research, there is a huge amount of data generated. In your production environment, there would not likely be as much. Still, making sense of sinkhole data can be tedious, so we connected our sinkhole to the McAfee® Enterprise Security Manager SIEM. In a single 24-hour period, we identified 96 unique malicious IPs that were sinkholed based upon our collection of more than 100,000 malicious IPs, domains and URLs, and more than 1 million events that passed through our sinkhole. We were also able to see the time distribution of attacks.

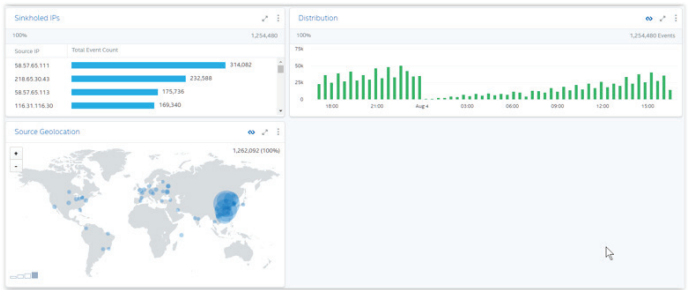

Figure 4. Top level view of sinkholed addresses.

We see some interesting data here. First, there are several IPs that seem to have a lot of activity associated with them. Second, we see a time distribution that is heaviest between around 3:00 pm Eastern time to about midnight. Finally, we see a strong cluster of sources in China. These findings are consistent with what Maltrail told us but are more detailed and granular.

If we look at remote access login more closely, which would be of considerable interest, we note that there were 2,183 attempts, all of which (plus others) were rejected.

Making sense of sinkhole data can be tedious, so we connected our sinkhole to the McAfee® Enterprise Security Manager SIEM. In a single 24-hour period, we identified 96 unique malicious IPs that were sinkholed based upon our collection of more than 100,000 malicious IPs, domains and URLs, and more than 1 million events that passed through our sinkhole.

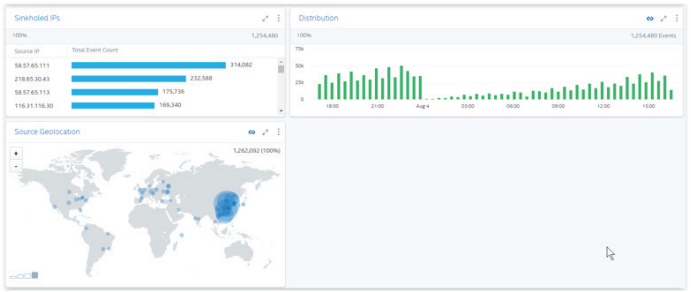

Figure 5. Sinkholed addresses detail.

We also see that there were mostly SSH/SSHD attempts, and they were thwarted. Finally, we see that most of the heavy hitters are from five cities in China. Drilling down to source geo, Jinan, we get three IPs: 58.57.65.111, 58.57.65.113 and 27.214.203.1. These three addresses account for nearly 480,000 attacks or scans. Now for more detail on each of these addresses. For example, have they been reported by others? We have some capabilities that we have added to our SIEM based upon its ability to execute certain types of commands, such as invoking the details from a threat intelligence website. We simply click on the address and select the function that we have added, in this case, CyMon (https://cymon.io/58.57.65.111 for the first IP). The result is that we see SSH attacks reported between April and end of July 2017. We also discover that it has been reported by the Alien Vault Open Threat Exchange. Again, we see reports of SSH attacks.

A Typical Threat Hunt

The SIEM is the hub of our threat hunting. From the SIEM, we get alerts that are a product of the data we feed to it from our on-network devices such as firewalls, data from open and closed source threat feeds, intelligence feeds, vulnerability assessments and threat calculations that are a product of threats, vulnerabilities, and weighting factors in the SIEM. For our threat hunt, we begin with an alert generated by a large number of events originating at IP 58.57.65.113. The events are directed at one of our external IPs. See Figure 6.

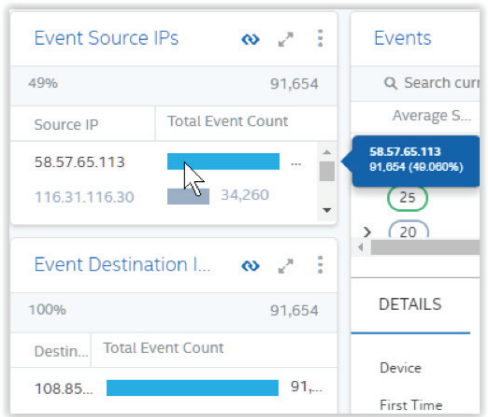

Figure 6. Large number of events from a single source IP.

The SIEM is the hub of our threat hunting. From the SIEM, we get alerts that are a product of the data we feed to it .

Our first task is to learn what we can about this IP. On our SIEM, we have added enrichment from several external sources, some open source, some not. See Figure 7.

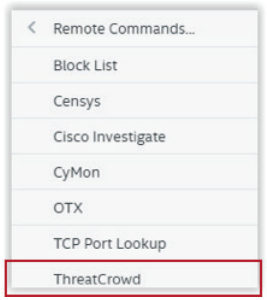

Figure 7. Remote commands programmed into our SIEM.

Our SIEM lets us add remote commands so, as you can see in the figure, we have added some threat and intelligence feeds. This was easy to do, since all we needed was to create a simple query that goes to a URL—the URL of the feed—and our SIEM will automatically append the IP address with which we are concerned to the URL. Now we can get data directly instead of having to go out to the feed separately and giving it our IP of interest. We can do this with a couple of mouse clicks. We begin with ThreatCrowd, an open-source service that gives us a lot of information about the IP in question. We click on ThreatCrowd, and we get the information in Figure 8.

Our SIEM also populates several threat intelligence watch lists from external open source and commercial threat sources. All events are then evaluated against these lists of malicious IP addresses, domains, and file hashes to determine additional situational awareness.

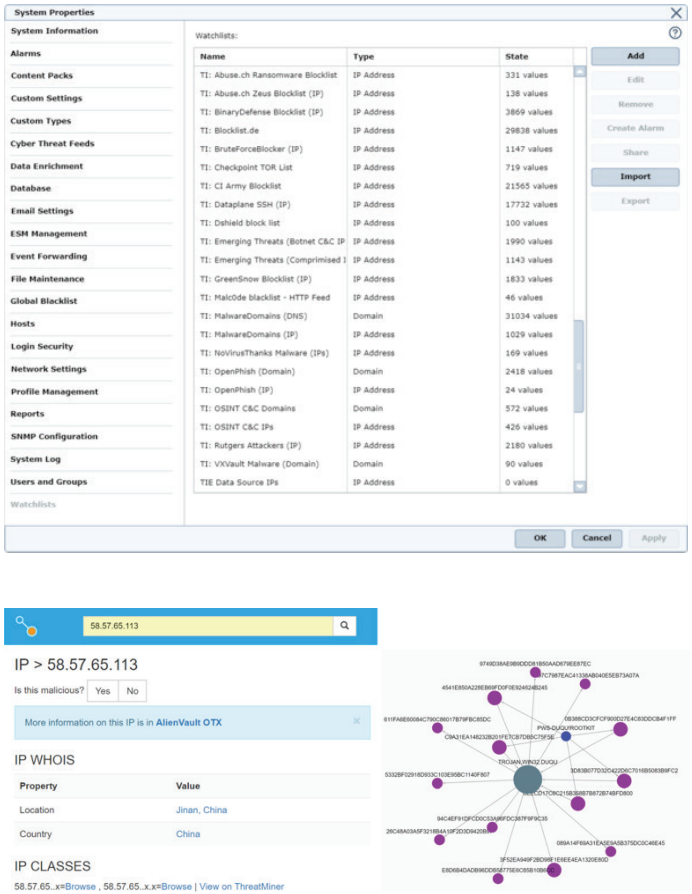

Figure 8. IP information from ThreatCrowd.

We notice a couple of important bits of information here. Our goal is to use the results of our threat hunt to inform our defense measures. So, we are quite interested in a couple of things. First, this is a state actor, or, at least, our actor is using a state internet service provider (ISP). We may assume, for now, that our actor is in China. The second thing that we note is that this IP does not seem to host any malicious domains. However, we do note that it is hosting three malware samples. These are very common samples, especially in China. For good measure, we can easily incorporate these file hashes into our SIEM watch list called "Known Malicious Hashes." Any future events containing these hashes will automatically be raised for inspection.

Doing a Google search on the IP by clicking the "Search in Google" button in Investigate, we see, among other things, an entry from the **Abuse IP Database** that tells us this IP has been reported frequently for SSH brute-force attacks. We can verify that this is what we are seeing by drilling down on the IP. It tells us, as noted in Figure 9, that port 22 was the target virtually all the time. Checking our logs, we find that 58.57.65.113 is a very frequent visitor. In fact, this IP performed the same type of brute-force attack against all the devices on our perimeter. We also see that it performed the attack over 4,300 times in a single 24-hour period.

This tells us several useful things. First, it tells us that the attack is automated—4,300 times in a single 24-hour period is an average of about three times per minute, which is too fast for a human, especially over a protracted period with no time gaps. Second, it tells us that the attempts are very persistent. We know from other intelligence sources that this type of automated, untargeted, attack is very typical of Chinese actors. We also see from our logs that the target port is always 22. That means that this specific attack is strictly SSH.

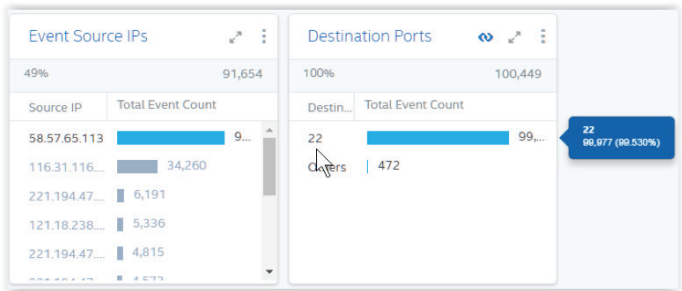

Figure 9. Port 22 is the target.

Now we have two things of importance from our threat hunt. First, we see that port 22 on our target machine has attracted attention. Second, we see that there is malware on the attacking site. Next, we want to see what other threat intelligence might be available to help us craft a strategy. One question we might have is whether or not this is a likely state actor or simply a hacker or hacker group operating within China. Evidence reported by foreignpolicy.com suggests that Chinese state hackers start their workday in the morning, Beijing time, and work into the afternoon, taking certain Chinese

Doing a Google search on the IP by clicking the "Search in Google" button in Investigate, we see, among other things, an entry from the Abuse IP Database that tells us this IP has been reported frequently for SSH brute-force attacks.

holidays off. Going back to our SIEM, we can look at the IPs passing through our sinkhole. We host a sinkhole and feed its output to the SIEM and chart the distribution across time. Taking a single 24-hour period, we get the information in Figure 10.

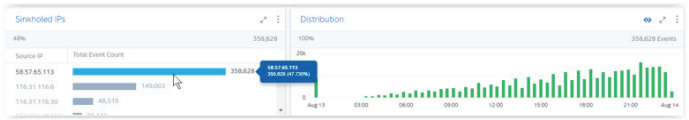

Figure 10. Time distribution for a 24-hour period of attacks from 58.57.65.113.

Note that the attack distribution starts at around 1500 Eastern time—about 0300 Beijing time—and starts to tail off at around midnight Eastern time, or noon Beijing time. We can draw some tentative conclusions. Either this is a private actor within China, a Chinese server has been co-opted by a non-Chinese actor, or the whole process is so automated that we cannot draw a conclusion. The third choice is the most likely since we do know, with a degree of certainty, that this is an automated process. That said, our actor could set his scripts to do whatever he wants and get a time distribution that is deceptive. In any event, this is a determined attack, and we need to take action.

We could, of course, simply shut down SSH on our target machine, but suppose that we have a business need to keep it open. Then what? The best bet in that case is to use two-factor authentication and monitor port 22 across all the outward-facing devices on our enterprise. In addition, our threat hunt suggested that there was malware—at least three unique samples—hosted on the attacker server. Given the hashes of those malware, we can arm ourselves to ensure that they are not in our enterprise and do not enter in the future. Finally, we know enough about this source IP to block it.

Our final step would be to perform an internal threat hunt for artifacts associated with 58.57.65.113 or, for that matter, any Chinese IP address, as well as artifacts, including quarantines, for the three types of malware found on the attacker's server. That threat hunt will be much like the external threat hunt we just completed, but we will focus upon internal IPs instead of the internet-facing ones. The one addition will be to employ the malware content pack, the PhishMe content pack, the recon content pack, the exploit content pack and, finally, the exfiltration content pack. Each of these will provide current state information about our enterprise and, moving backward in time, a historical view as well.

Summary

If we think of a threat hunt as a wheel with each of our threat hunting tools as spokes, the SIEM is the hub. Our tools may be external feeds, feeds from our internal sensors, or the feeds included with the SIEM itself. To understand asset-based risk, we need to add vulnerability assessment and asset weighting. However, risk, while very important, is not, by itself, a key aspect of threat hunting.

If we think of a threat hunt as a wheel with each of our threat hunting tools as spokes, the SIEM is the hub.

Strategy, tactics, and operations

The military view is useful here. We are concerned with strategic, tactical and operational aspects of our defenses. Risk is strategic, threat hunting is tactical, and remediation is operational. We also should view the progress of an attack. Attacks begin with preparing the battle space. That is exactly what we saw in our SSH example. The attacker, without focusing on a specific exploit, is scanning/brute-forcing a service known to be potentially vulnerable. Should the brute-forcing efforts prove successful, they would be followed up by a specific campaign to exploit the vulnerabilities reachable from the SSH service.

We can see an example of a campaign using a Structured Threat Information Expression (STIX) campaign-level view of a campaign by an actor called japanorus in Figure 11.

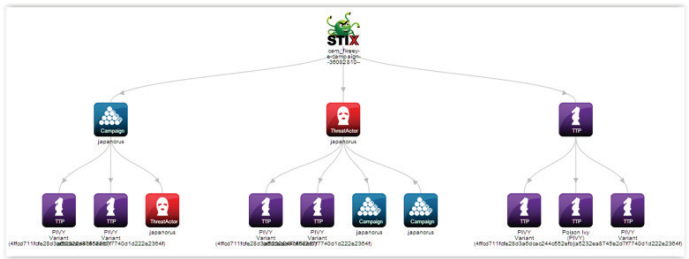

Figure 11. STIX view of a japanorus campaign.

In this case, the STIX view shows the actor, the campaign, and the tactics, techniques, and procedures (TTPs). A broader view, such as in Figure 12, adds indicators and observables among other STIX elements. However, in Figure 11, we can see that the actor uses variants of the Poison Ivy malware. Poison Ivy is a remote administration tool (RAT) and is used frequently as a back door. It is old, but it also is very customizable, so it is found frequently even today. That knowledge gives us something to look for. It also helps us understand how an attack may progress.

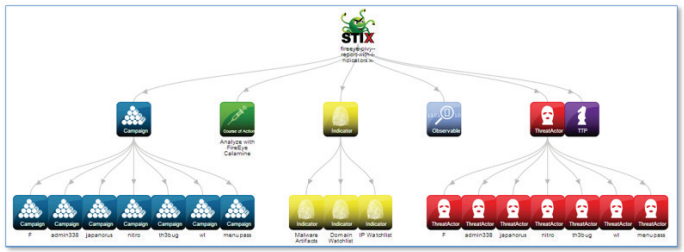

Figure 12. A collection of actors and campaigns using Poison Ivy as the primary TTP.

As you can see in Figure 12, we have added several campaigns to the japanorus campaign, all of which have the use of Poison Ivy as a TTP in common. Here we see campaigns, indicators, actors, observable, TTPs, and the green course of action. Each of these can be expanded but would be undecipherable on our illustration. By expanding based upon a known TTP, we get a view of the campaigns that currently are in progress that use the specific TTP—possibly among others—and we can prepare for them. Preparing the battle space goes both ways. The attacker will perform reconnaissance and pre-scanning, and we must anticipate those activities and prepare the battle space to reject the scans.

An effective SIEM will also support the consumption of indicator of compromise (IoC) content to enable more advanced and automated threat hunting. Our SIEM consumes from a Trusted Automated eXchange of Indicator Information (TAXII) source or directly from a sandbox conviction and performs an historical "backtrace," identifying any previous events containing one or more of the threat artifacts/observables listed in the IoC threat feed. More advanced hunting tools, such as endpoint detection and remediation (EDR), can also be integrated into an effective SIEM threat research practice, allowing SOC analysts to search in real time for artifacts residing on endpoints and automatically eliminate unwanted and potentially malicious files, registry values, and applications.

The intelligent use of an intelligent SIEM is the key to managing the strategic, tactical and operational aspects of threat hunting. In today's threatscape, we cannot ignore any of the three. Effective integration of SIEM as the hub and an arsenal of threat investigation tools as the spokes is critical to gaining enhanced visibility of the hazards coming down the road. And seeing the threat landscape clearly is a business imperative that demands close attention.

1. https://en.wikipedia.org/wiki/Cyber_threat_hunting#cite_note-1
2. http://foreignpolicy.com/2013/01/31/the-peoples-republic-of-hacking/

About McAfee

McAfee is one of the world's leading independent cybersecurity companies. Inspired by the power of working together, McAfee creates business and consumer solutions that make the world a safer place. By building solutions that work with other companies' products, McAfee helps businesses orchestrate cyber environments that are truly integrated, where protection, detection and correction of threats happen simultaneously and collaboratively. By protecting consumers across all their devices, McAfee secures their digital lifestyle at home and away. By working with other security players, McAfee is leading the effort to unite against cybercriminals for the benefit of all.

www.mcafee.com.

2821 Mission College Boulevard
Santa Clara, CA 95054
888 847 8766
www.mcafee.com