Report

2012 Threats Predictions

By McAfee® Labs™

Table of Contents

Predicting future threats can be a hit-or-miss exercise for a security research organization. Certainly it is interesting to put on our wizard hats and prognosticate about what may happen in the coming months, but how much do threats really change each year? The past 12 months were a transformative year in many ways, but were these transformations revolutionary or evolutionary? We saw great changes in mobile threats, hacktivism, client-side exploitation, social-media exploitation, and targeted attacks. Many of these changes and trends will continue to influence the threats landscape for years to come.

What changes to threats does McAfee Labs expect in the coming year? We foresee several new scenarios as well as some significant evolutions in even the most established threat vectors:

- Industrial threats will mature and segment
- Embedded hardware attacks will widen and deepen
- Hacktivism and Anonymous will reboot and evolve
- Virtual currency systems will experience broader and more frequent attacks
- This will be the "Year *for* (not "of") Cyberwar"
- DNSSEC will drive new network threat vectors
- Traditional spam will go "legit," while spearphishing will evolve into the targeted messaging attack
- Mobile botnets and rootkits will mature and converge
- Rogue certificates and rogue certificate authorities will undermine users' confidence
- Advances in operating systems and security will drive next-generation botnets and rootkits

The stage is set, so let's move on to the specifics!

Industrial Threats

Threats to industrial and national infrastructure networks have recently garnered a lot of attention, and there is a very good reason for that. This is one of the few areas in which a cyberthreat endangers the real loss of property and life. Industrial SCADA (supervisory control and data acquisition) systems are just as vulnerable as any other networked system, but the big difference is that many these systems were not designed for the networked environment the world continues to adopt. Increased interconnectivity for systems and devices not designed for this type of access is a recipe for trouble—due to the lack of information security practices in many of the environments SCADA systems are deployed in. It seems to be a common practice to connect critical infrastructure systems to the Internet and then manage them with commonly available software. All software has vulnerabilities, but industrial IT systems require greater diligence in architecture, design, and implementation. Attackers will leverage this lack of preparedness with greater frequency and success in 2012, if only for blackmail or extortion. When one considers the goals of many hacktivist groups, the possible mating of political goals with vulnerabilities in industrial controller systems (ICS) needs to be taken *very* seriously.

Stuxnet proves that malicious code can create a real world, kinetic response.[1] Recent incidents directed at water utilities in the United States show that these facilities are of increasing interest to attackers. The more attention is focused on SCADA and infrastructure systems, the more insecurity seems to come to light. We expect to see this insecurity lead to greater threats through exploit toolkits and frameworks as well as the increased targeting of utilities and energy ICS systems in particular. Once a targeted group has been shown to have a soft center, the attackers will dig in eagerly.

Attackers tend to go after systems that can be successfully compromised, and ICS systems have shown themselves to be a target-rich environment. Their administrators should take heed of recent events. It's time for extensive penetration testing and emergency response planning that includes cybercomponents and networking with law enforcement at all levels. They must ask themselves: What happens when we are targeted?

The Threat Within: Embedded Hardware

Embedded systems have grown in popularity and importance during the last several years. In general, these are designed for a specific control function within a larger system, often with real-time computing requirements. They often reside within a complete device that includes hardware and other mechanical parts. Historically used for industrial needs such as avionics, transportation, and energy as well as automotive and medical devices, this architecture is increasingly making its way into the business, enterprise, and consumer worlds. GPS, routers, network bridges, and recently many consumer electronic devices use embedded functions and designs.

Exploiting embedded systems will require malware that attacks at the hardware layer; that type of expertise has ramifications that go beyond embedded platforms.

Malware writers now create malware that targets the lower parts of the operating system more and more often. Many times attackers will try to "root" a system at its lowest level, including the master boot record and even BIOS layers. If attackers can insert code that alters the boot order or loading order of the operating system, they will gain greater control and can maintain long-term access to the system and its data. Controlling hardware is the promised land of sophisticated attackers.

The consequence of this trend is that other systems that use embedded hardware will become susceptible to these types of attacks. We have seen concept code that targets the embedded hardware in automotive systems, medical systems, and utility systems. We expect these proofs-of-concept code to become more effective in 2012 and beyond.

Hacktivism

Although hacktivism is not new, with the WikiLeaks saga on the front pages in 2010 hacktivism gained wider publicity, acceptance, and usage than ever before. Overall, 2011 was a muddled year for online activists, with conflicting players frequently at odds with each other and no clearly stated goals. It was often difficult to sort things out between politically motivated campaigns and simple script-kiddies entertainment, but one thing became clear: When hacktivists picked a target, that target was compromised either through a data breach or denial of service. They are a credible force. Agree with their goals or not, Anonymous and other hacktivist groups have shown themselves to be dedicated, resourceful, and even agile in choosing some of their targets and operations.

The coming year will be decisive for hacktivism. And the Anonymous stories represent only one aspect of this issue.

- The "true" Anonymous (that is, its historical wing) will reinvent themselves and their scene or die out. If the Anonymous circles of influence are unable to become organized—with clear calls for action and responsibility claims—all those labeling themselves Anonymous will eventually run the risk of becoming marginalized. Either way, we will see a large increase in such attacks. Distributed denial of service (DDoS) and personal data disclosures justified by a political conscience will continue to grow.
- The people leading digital disruptions will become better engaged with the people leading physical demonstrations. We will see more mating of social media-based hacktivism with social media-coordinated hacktivism. We expect many future operations to include both physical and digital components. Joint and coordinated actions, in the field and online, will be simultaneously planned. It is not hard to predict the evolution of the Occupy and other outraged groups to include more direct digital actions. As we posited in other predictions, the possibility of mating hacktivist goals with industrial controller or SCADA system availability is a very real possibility. We expect hard-line hacktivists supporting the worldwide Occupy movements will drop the Anonymous label and soon operate as "Cyberoccupiers."
- For political and ideological ends, the private lives of public figures—politicians, industry leaders, judges, and law-enforcement and security officers—will be disclosed this year more than in the past. Protesters will stop at nothing to obtain data from social networks or web servers to support their various operations.

- Some hacktivists will operate along the same lines as the various "cyberarmies" that primarily flourish in nondemocratic or nonsecular states (Iranian Cyber Army, Pakistan Cyber Army, ChinaHonker group, etc.). Mostly used for defacement in the past two years, the armies will move to more disruptive actions in the new year. Some of these groups will clash themselves, possibly causing unpredictable collateral damages (Palestinian versus Israeli, Indian versus Pakistani, North versus South Korean, etc.). In 2011, cyberarmies were rumored to be manipulated or supported by their governments. Totalitarian states will go further next year, even acknowledging the actions of local cyberarmies.

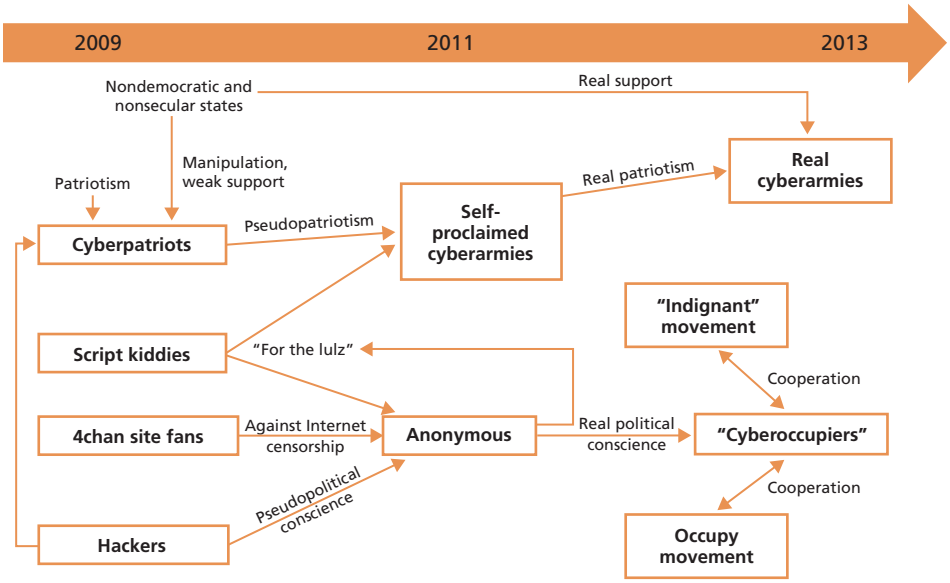

Figure 1. The many connections and motivations of hacktivism.

Virtual Currency

Virtual currency, sometimes called cybercurrency, has become a popular way for people to exchange money online. Though not necessarily backed by tangible assets or even tangible commodities, services such as Bitcoin allow users to make transactions through a decentralized, peer-to-peer network—essentially electronic cash that allows direct, online payments. A user needs only client software and an online wallet service to receive the "coins," which are stored in the wallet and can be transferred to others as payment for goods or services. For users to send or receive these coins, they simply need a wallet address. Can you see both the problem and the opportunity?

Trojan malware easily fits into this architecture. The wallets are not encrypted and the transactions are public. This makes an attractive target for cybercriminals. Several events of note took place in 2011 regarding virtual currencies:

- The Mt. Gox Bitcoin Exchange database was targeted by attackers who stole thousands of Bitcoins
- Spam promoting fake Bitcoin mining tools was distributed. These tools actually contained malware designed to send the victims' wallet files to a remote location. It also allowed other miners to use the infected computer for further Bitcoin mining.
- Bitcoin miner botnets were found in the wild. Using large numbers of infected machines, these botnets could speed up Bitcoin mining and processing and could also launch DDoS attacks.

The nature of virtual currencies and technologies like Bitcoin are too good a target for cybercriminals to pass up. We saw considerable growth in malware that targets these technologies in 2011. Here is a look at Bitcoin malware in particular:

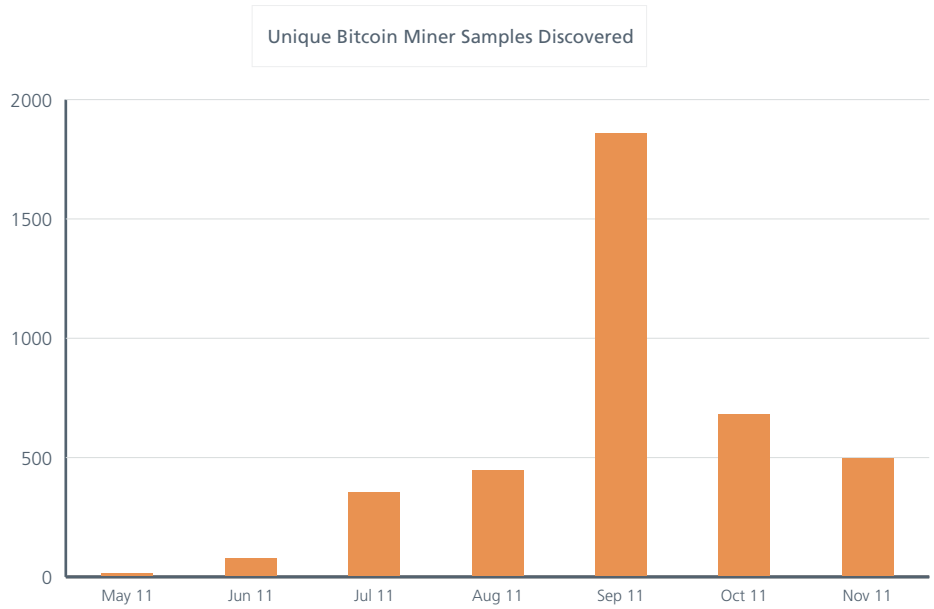

Figure 2. The theft (called mining) of the virtual currency Bitcoin reached a peak in September. We predict the rip-offs will increase in 2012.

We expect to see this threat evolve into a cottage industry of cybercrime next year—with spam, data theft, tools, support networks, and other associated services dedicated solely to exploiting virtual currencies. Clearly, cybercriminals have found a payment system that fits their needs.

Cyberwar

Will this be the Year of Cyberwar, or merely a showcase of offensive cyberweapons and their potential? While we certainly hope it's only the latter, the situation's growth during recent years makes an eventual cyberwar nearly inevitable. We have frequently seen "cyber" techniques complement traditional methods of intelligence, or espionage, operations, with many players accusing others, friends and foes alike. It's a very cheap way of spying, always leaves room for plausible deniability, doesn't endanger human lives and, most important, seems to be highly effective. What we haven't much seen is the use of cyber as part of the arsenal in an armed conflict. So far this has been witnessed only on a rather small scale with very limited sophistication of the attacks, for example, in the Georgia conflict.

But now the situation has changed. Many countries realize the crippling potential of cyberattacks against critical infrastructure and how difficult it is to defend against them. Their potential opens up opportunities for attack by small countries or organizations, particularly if there are few targets to strike back against. The Stuxnet attack was a game-changing event in many aspects; one of them was to make it absolutely clear to everyone that the threat is real and what impact such attacks could have.

The United States realizes how vulnerable it is, probably more than any other country because of its massive dependence on computer systems and a cyberdefense that pretty much defends only government and military networks (imagine an army that protects only military bases rather than any other part of the country). After taking a lot of criticism for the absence of a formal doctrine, the country finally reacted.

In July the "Department of Defense Strategy for Operating in Cyberspace" was released.[2] The report states "Strategic Initiative 1: DoD will treat cyberspace as an operational domain to organize, train, and equip so that DoD can take full advantage of cyberspace's potential." But you won't find in this paper a topic that was discussed previously—that cyberattacks of sufficient impact could meet with a return strike. Instead the DoD is preparing a new doctrine to complement the cyberstrategy that offers concrete guidance for the DoD's cyberwarfare staff. If that doctrine outlines under what circumstances a cyberretaliation can be considered, this would still be a far step from the "threat of total annihilation" doctrine that helped the world survive the cold war.

It doesn't really deter anyone from attacking if the possible response is unknown because it's classified.

According to reports, the use of cyberweapons in the revolution in Libya was considered but didn't happen because no one wanted to be the first to open Pandora's box. Or maybe it just wasn't a target-rich environment. For now, however, we've seen no public demonstration of the capabilities of offensive cyberwarfare that have the potential to deter anyone. The voices are getting louder to declassify that information, so some sort of demonstration—other than showing frightening videos of failing machinery to foreign diplomats—is expected. An effective demonstration has the potential to trigger a "me too" response from other states, showing that they also have the same capabilities.

We hope in the coming year to see only demonstrations, rather than any effects of an actual cyberwar!

DNSSEC

DNSSEC (Domain Name System Security Extensions) is a technology to protect name-resolution services from spoofing and cache poisoning by using a "web of trust" based on public-key cryptography. This is meant to protect a client computer from inadvertently communicating with a host as a result of a man-in-the-middle attack, which redirects the traffic from the intended server (web page, email, etc.) to another server. To protect online users and implement a more difficult terrain for hackers, this is an extremely important step in the evolution of the Internet.

Unfortunately DNSSEC would also protect from spoofing and redirection any attempts by authorities who seek to reroute Internet traffic destined to websites that are trafficking in illegal software or images. For a government to redirect traffic, it would need to be considered authoritative from the root-level domains, which is a level of trust that other governing bodies would hesitate to grant if they knew that the result would be the suppression of Internet content based on the opinions of foreign governments.

Recent attempts to pass legislation to prevent the disbursement of intellectual property are based on an understanding of the current state of how today's DNS works and not how the future DNSSEC will work. This gap may create additional legal requirements for managing current DNS infrastructure, which may not be compatible with DNSSEC infrastructure. If such requirements are implemented, then the process of upgrading the security of our DNS infrastructure may be put on hold while committees seek a technical middle ground between the law and DNSSEC.

With governing bodies around the globe taking a greater interest in establishing "rules of the road" for Internet traffic, we can expect to see more and more instances in which the solutions of tomorrow are hampered by legislative wrangling over the issues of yesterday. The result is that the Internet of tomorrow will probably look like the Internet of yesterday for a longer time than we security folks would like to see.

Spam Goes "Legit"

During the past four years we've seen increased international understanding and cooperation in combating botnet-related spam. This cooperation has resulted in a number of high-profile shutdowns of infrastructure that was critical to botnet control (such as the ISP McColo), spam domain webhosting (Glavmed), credit card processing linked to counterfeit pharmaceuticals, and even suits against large Internet corporations that provided advertising outlets for illegal enterprises. These actions have resulted in an enormous drop in global spam volumes from a peak in mid-2009 and significantly increased the black-market cost of sending spam through botnets.

Although these steps by no means represent the end of all spam—as some technology prophets have predicted—they do change the landscape. Today as we look across that landscape, we see more and more unsolicited spam mail being sent not from botnet-infected hosts, but by actual "legitimate" advertising agencies that use techniques heavily derided by the antispam community. Their efforts result in users' email addresses getting on advertising lists without their knowledge or consent. These techniques range from blatant purchasing of email address lists that are advertised as offering users who have already consented to receive any advertising (a claim that requires a willing suspension of disbelief), to "e-pending" (harvesting email addresses through algorithms that determine that people would sign up for advertising if they were offered the chance, then skipping the asking part and just adding them to a list without permission), to purchasing customer databases from companies going out of business and ignoring any privacy policy that was in place when the company was still operating, to "partnering" with other advertising entities or mailing-list providers to blitz their email lists with advertising.

The advertising companies that do this know that they're sending spam and use the same techniques that botnet operators use to attempt to evade detection. Every day thousands of new email domains are registered using whois privacy to prevent identification of the owner, and thousands of new IP addresses are activated in the subnets of hosting providers for a few hours of a spam cannon that plasters inboxes with poorly formatted emails, fraught with misspellings and bad grammar. Most of these emails contain an opt-out link that doesn't accomplish anything except to let the spammers know that your email address is active and you're reading their mail. And there is an address where you can send a snail mail to get delisted (but if you look up the addresses online they can range from shacks in the middle of the Canadian wilderness to barren plots of land in the Arizona desert). In some cases individual email addresses have received more than 9,000 nearly identical spam messages in one day advertising the health benefits of a popular magnetic bracelet.

These corrupt advertising practices are supported by law. The United States' CAN-SPAM Act was watered down so much that advertisers are not required to receive consent for sending advertising. Because advertising is such a profitable business, with plenty of lobbying prowess, it is extremely unlikely that any significant changes to email list-management practices or large penalties for bad behavior are anywhere on the horizon.

In this environment, we can expect to see "legal" spam continue to grow at an alarming rate. It is cheaper and less risky to spam individuals from advertising companies than it is to use botnet-infected hosts. This sort of activity, known as snowshoe spamming, has grown so much that at the time of this writing the top 10 most common email subjects include one delivery status notification, one botnet-related fake-Rolex spam, one confidence scam, and seven subjects associated with snowshoe spam. This sort of traffic will continue to grow at a faster rate than phishing and confidence scams, while botnet-related spam will continue to decrease as botmasters find better and safer ways to wring money out of their armies of infected computers. It is only a matter of time before most global spam volume comes from badly behaving but "legal" entities.

Mobile Threats

During the last two years we've seen an increase in attacks on smartphones and mobile devices. We've run across rootkits, botnets, and other malware. Attackers have moved on from simple destructive malware to spyware and malware that makes them money. We've seen them exploit vulnerabilities to bypass system protections and gain greater control over mobile devices. In 2012 we expect to see attackers continue what they've done and to improve upon their attacks. We also predict a move toward mobile-banking attacks.

Botnets + rootkits = low-level trouble

On PCs, rootkits and botnets deliver ads and make money off of their victims. On mobiles, we've seen these types of malware used in the same manner. Rootkits allow the installation of additional software or spyware, and botnets can cause ad clicks or send premium-rate text messages.

We've seen mobile variants of malware families that include Android/DrdDream, Android/DrdDreamLite, and Android/Geinimi, as well as Android/Toplank and Android/DroidKungFu. Some of these malware have used root exploits, originally developed for customers to unlock their own phones, to gain access and take over victims' phones. In the coming year as developers and researchers develop new methods for rooting phones, we will see malware authors adapting the lessons of PC malware development to undertake attacks that leverage the mobile hardware layer to a greater extent. PC-based malware is increasingly moving further "down" the operating system (OS) to take greater advantage of hardware; we expect mobile malware to follow the same direction.

Bootkits, malware that replaces or bypasses system startup, also threatens mobile devices. Although rooting one's own phone or ebook reader opens the device to extra features or to replacing the OS, it can also allow attackers to load their own modified OS. Whereas a mobile rootkit will simply modify the existing OS to evade detection, a bootkit can give an attacker much greater control over a device.

For example, the "Weapon of Mass Destruction" mobile penetration-testing toolkit runs on old Windows Mobile phones. WMD installs itself using tools developed to load Linux on Windows Mobile phones and allows the user to reboot to the original OS. Attackers have already used old root exploits to hide themselves; as new exploits are developed, attackers will eventually install their own custom firmware.

Mobile banking attacks

PC users have seen attacks from criminals using the Zeus and SpyEye crimeware kits to steal money from online banking accounts. Both Zeus and SpyEye have begun to use mobile apps as helpers to bypass two-factor authentication and gain access to victims' money.

Zitmo (Zeus-in-the-mobile) and Spitmo (SpyEye-in-the-mobile) are two families of mobile spyware that forward SMS messages to attackers. Using this spyware required the attackers to log in manually to steal users' money.

Last July, security researcher Ryan Sherstobitoff discussed how the transactions performed by criminals using Zeus and SpyEye could be tracked—as they looked nothing like those of legitimate users. Last month, he showed how criminals had adapted and now can programmatically steal from victims while they are still logged on. This helps the criminals transactions appear to come from the legitimate users and by adding a delay seem to be performed by a real human. Attackers have adapted quickly to every change intended to secure banking on PCs. As we use our mobile devices ever more for banking, we will see attackers bypass PCs and go straight after mobile-banking apps. We expect to see attacks that leverage this type of programmatic technique in greater frequency as more and more users handle their finances on mobile devices.

Rogue Certificates

We tend to believe in files and documents when they are digitally signed due to our trust in digital signatures and the certificate authorities they come from. Many whitelisting and application control systems depend on valid digital signatures. These solutions allow us to put policies and controls in place around services, applications, and even files that carry a valid digital signature. Secure web browsing and secure online business transactions also rely on trusted digital signatures. These certificate authorities and their "certs" basically tell the operating system "You can trust me because I am valid and vouched for."

Given that trust, what happens if we're faced with rogue or fake digital certificates? Going deeper, what are the implications of a certificate authority that is compromised? Digital certificates allow us a certain level of trust in a file, process, or transaction. By producing and circulating fake or rogue certificates, attackers can engage in almost undetectable attacks. On the browser, this allows an attacker to engage in man-in-the-middle attacks: traffic that was otherwise encrypted and not viewable to the attacker can now be seen plain as clear text because they have the "key." On the host, security software will ignore a file signed with a valid key as it now appears to be whitelisted: It has authorized access due to the certificate it presents.

Recent threats such as Stuxnet and Duqu used rogue certificates to great effect to evade detection. Although this is not the first time we have seen this behavior (fake AV, certain Zeus variants, Conficker, and even some old Symbian malware used them), we expect to see this trend increase in 2012 and beyond.

The larger threat of targeting certificate authorities to produce rogue certificates is also a concern for the future because this type of compromise would allow an attacker to create multiple keys to be used in a variety of web-based and host-based scenarios, effectively undermining much of the trust that is built into an operating system. We are very concerned about the implications of large-scale rogue certificates on the whitelisting and application control technologies that use these certs. DigiNotar, an already troubled Dutch authority, recently declared bankruptcy after a security breach resulted in the issuance of fraudulent certificates. Was this attack the final nail in its coffin? Investigations have shown that as many as 531 fraudulent certificates were issued from DigiNotar. It is probable that the company's fall is only the beginning of our insight into breaches in this industry. Now we must worry about how much trust and damage has been done.

Wide-scale targeting of certificate authorities and the broader use of fraudulent, yet valid digital certificates has ramifications for public-key infrastructure, secure browsing, and transactions as well as host-based technologies such as whitelisting and application control. Taking advantage of our trust in this system gives attackers a great advantage; we certainly expect them to focus on this area.

Advances in Operating Systems

Information security always involves give and take, with equal amounts of measures and countermeasures thrown in. The attackers write malicious code; we counter it. Operating system vendors bake security into the core of the OS; attackers find a way to circumvent. This is a natural part of the dynamic threat landscape and will never go away. But will advances by the information security industry and operating system vendors drive malware writers outside the OS to directly attack hardware?

Recent versions of Windows have included data-execution protection as well as address-space layout randomization. These security methods make it harder for attackers to compromise a victim's machine. Encryption technologies have also boosted OS protection in recent years. As with most internal OS security measures, attackers very quickly found ways to evade them. With the upcoming release of Windows 8, Microsoft will include many new security features: secure password storage, secure boot functions, antimalware defenses, and even enhanced reputation capabilities. Where will this new security architecture drive attackers?

The answer is "down and out": down into hardware and out of the operating system.

During the last several years McAfee Labs has seen great advances from attackers and malware writers in both rootkits and bootkits. Rookits are used to subvert both the operating system and security software, while bootkits attack encryption and can replace legitimate boot loaders. These are advanced techniques to intercept encryption keys and passwords, and even subvert driver-signing defenses employed by some OS's.

Attacking hardware and firmware is not easy, but success there would allow attackers to create persistent malware "images" in network cards, hard drives, and even system BIOS. We expect to see more effort put into hardware and firmware exploits and their related real-world attacks throughout 2012 and beyond.

Advances in the Windows 8 bootloader security feature have already caused researchers to show how they can be subverted through legacy BIOS; meanwhile, the product has not even been fully released yet. With further development around Intel's unified extensible firmware interface specifications—designed as a software interface between the operating system and platform firmware to enforce a secure boot and to replace legacy BIOS—we expect more attackers to devote their time to evasion research in the coming years.

We will keenly watch how attackers use these low-level functions for botnet control, perhaps migrating their control functions into graphics processor functions, the BIOS, or the master boot record. At the same time we expect attackers to leverage "new" protocols standards such as IPv6 as network implementations advance along the lines of operating systems.

In spite of our efforts to thwart their ambitions, attackers clearly see the value and power of attacking hardware and moving outside of tradition operating system attacks.

About the Authors

This report was prepared and written by Zheng Bu, Toralv Dirro, Paula Greve, David Marcus, François Paget, Ryan Permeh, Craig Schmugar, Jimmy Shah, Peter Szor, Guilherme Venere, and Adam Wosotowsky of McAfee Labs.

About McAfee Labs

McAfee Labs is the global research team of McAfee. With the only research organization devoted to all threat vectors—malware, web, email, network, and vulnerabilities—McAfee Labs gathers intelligence from its millions of sensors and its cloud-based service McAfee Global Threat Intelligence™. The McAfee Labs team of 350 multidisciplinary researchers in 30 countries follows the complete range of threats in real time, identifying application vulnerabilities, analyzing and correlating risks, and enabling instant remediation to protect enterprises and the public.

About McAfee

McAfee, a wholly owned subsidiary of Intel Corporation (NASDAQ:INTC), is the world's largest dedicated security technology company. McAfee delivers proactive and proven solutions and services that help secure systems, networks, and mobile devices around the world, allowing users to safely connect to the Internet, browse, and shop the web more securely. Backed by its unrivaled Global Threat Intelligence, McAfee creates innovative products that empower home users, businesses, the public sector, and service providers by enabling them to prove compliance with regulations, protect data, prevent disruptions, identify vulnerabilities, and continuously monitor and improve their security. McAfee is relentlessly focused on finding new ways to keep our customers safe. http://www.mcafee.com

[1] https://blogs.mcafee.com/mcafee-labs/stuxnet-update
[2] Read the unclassified version at http://www.defense.gov/news/d20110714cyber.pdf

McAfee
An Intel Company

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

40302rpt_threat-predictions_1211_fnl_ETMG