Are you feeling

Detecting and dealing with Advanced Persistent Threats to embedded systems. By **Avishai Ziv**.

As the vision of 15billion devices being connected to the Internet of Things by 2015 gets closer, major concerns are emerging about the security of the infrastructure that makes it all happen.

High profile targeted attacks through connected devices have highlighted the fact that, whilst industry is aware of the need to protect against common malware attacks, there is insufficient awareness of the need to deal with direct and targeted attacks on specific pieces of infrastructure.

When connected embedded devices use the same operating systems as IT endpoints (ATMs and POS terminals, for example), hackers can use well tried techniques to attack the embedded infrastructure. This is alarming when you consider that connected embedded devices are being used to control strategic infrastructure – the national grid for example. Welcome to the world of Advanced Persistent Threats (APTs).

The corporate IT world has been looking at the issue of APTs for some time, but with little success. While new solutions to tackle APTs are being introduced continuously, the detection gap remains alarmingly long. The main reason is that common security solutions fail to detect the actual APT infection. Instead, they focus on failed prevention attempts (using conventional anti malware technologies) and on monitoring targets that are already infected.

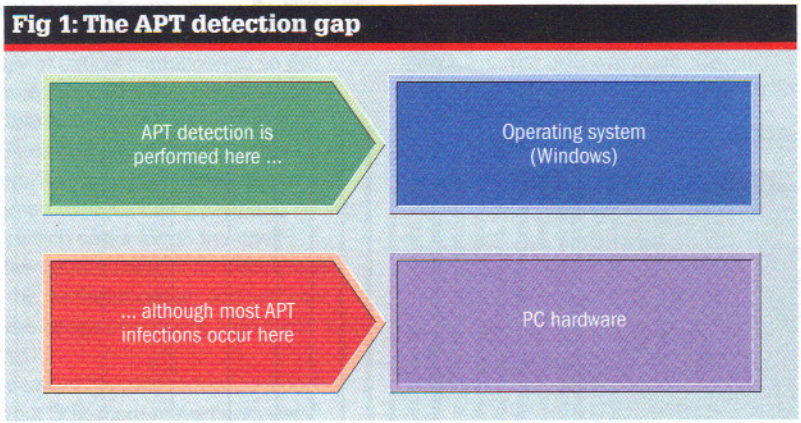
So the question remains: how can such attacks be detected effectively and averted in the embedded world, where timely detection is paramount? Using a new and unconventional method of detection – namely a secure embedded hypervisor – can resolve that problem.

While security vendors are responding more quickly to new methods of infection and evasion, it still takes months to detect APTs – the industry's accepted average is from 6 to 9 months.

The main reason for the APT detection gap is the sophistication of infection techniques used by the attackers. Most infections occur beneath the infected operating system and, as such, cannot be seen in real time by common detection technologies like anti-malware applications and sandboxes.

While security vendors are responding more quickly to new methods of infection and evasion, it still takes months to detect APTs – the industry's accepted average is from 6 to 9 months.

Fig 1: The APT detection gap



The main stages of APT attack

Before anything happens, the attackers will use reconnaissance techniques to identify the target, obtain user contacts and prepare for the attack. However, the APT attack begins only when it reaches the target – usually an endpoint.

An APT attack has three phases:

- **Penetration.** Exploiting vulnerabilities in an operating system and/or an application in order to allow the APT to be installed. This is a glaring weakness in the embedded world as embedded terminals typically run an outdated operating system (say Windows XP), which is updated infrequently, or not at all.
- **Infection.** Installation of the APT, commonly referred to as 'dropping', with the APT component (mostly with a rootkit module) known as the 'payload'. This is the critical stage at which the target is compromised and the APT gains enough control over the target machine to carry out its malicious tasks.
- **APT activity.** Malicious activity on the infected and compromised machine, including communication with the C&C server, gathering personal information, deleting data and so on.

A gaping hole

APTs are currently detected using:

- Common anti-malware products (client applications, gateways, sandboxes and cloud services), which try to detect and prevent penetration.
- Existing anti-APT solutions, which focus on the APT's activity in the infected machine by discovering and monitoring the APT's network activity (mostly outbound traffic). These solutions do not prevent infection, nor are they capable of detecting the

threatened?

infection prior to its network activity.

There is no solution capable of detecting the actual APT infection – the most critical and vital stage of APT attacks – and issue an alert when it happens. Hence, there is an APT detection gap.

Detection challenges

Most APTs use low level and sub OS rootkits, which are designed to be undetectable by the OS or any security application installed in or on it (and thus perform the 'P' – persistence – element of APT).

In order to be undetectable, yet gain enough control over vital OS functions, a rootkit typically needs:

- to install itself in parts of the hard disk hidden from view (unpartitioned sectors between the disk partitions and the last disk sector) and to access the OS.
- to obtain a security privilege superior to that of the infected OS (subverting the boot sequence of the OS and launching itself before the OS by altering OS original boot sectors – master boot record (MBR), volume boot record (VBR) and the Unified Extensible Firmware Interface (UEFI).

These two rootkit traits are deadly; while the infectors used to penetrate the OS change rapidly, rootkits change more slowly. Because they are stealthy and undetectable, new versions of these rootkits may only appear once every 12 to 18 months because there is little need to change.

Hypervisor honeypot

A new approach must: detect an APT infection in real time; and provide threat response personnel with live forensics data in order to cut their analysis and response time.

- **Detection:** Given the evasive nature of APTs, detection must be carried out at a level lower than their level of infection and activity. That can only be a bare metal hypervisor (such as LynxSecure), separating the hardware from the software, presenting only bare virtual hardware to the installed OS. Effectively a 'virtual motherboard', such a system will be invisible to APTs and undetectable by them.

The hypervisor must be designed specifically to serve as the means of detection – a honeypot – and hardened rigorously so that it will not become a target for those attacks. This approach also removes any OS dependency (a deficiency of some existing solutions), meaning detection becomes OS agnostic. As the most privileged monitor in the platform, it will be able to detect any changes to the monitored hardware.

A properly designed embedded hypervisor, with a very small inherently secure code base, can be installed on typical PC based embedded systems, as these tend to be very humble in terms of compute power and memory size.

A properly designed embedded hypervisor, with a small, inherently secure code base, can be installed on typical PC based embedded systems, as these tend to have restricted compute power and memory size. The hypervisor's small size will further strengthen the security of the system and reduce the attack surface.

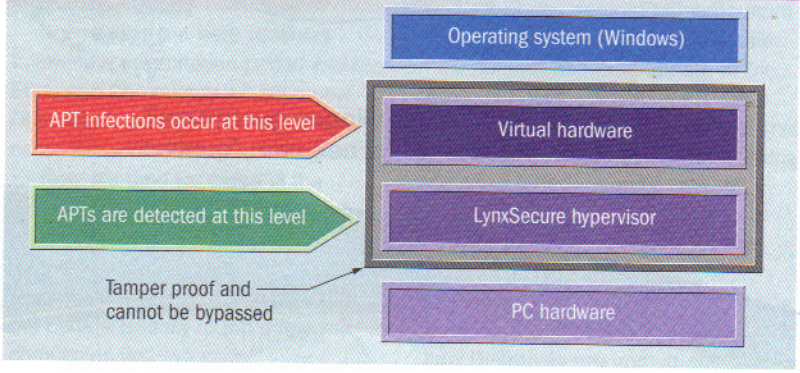
In this way, the stealthiest rootkits – such as MBR wipers (Dark Seoul), MBR infectors (TLD4), VBR infectors (XPAJ) and malware using hidden file systems (ZeroAccess) – will be intercepted immediately.

- **Live forensics.** Currently, finding the exact details of such infections requires arduous and lengthy forensic analysis; the forensics data of the uninfected hardware is not available and the entire hardware needs to be inspected. It is like looking for a needle in a haystack.

However, a secure embedded hypervisor allows the generation of immediate fine tuned forensics reports, containing only the infected sections, with an automated analysis of the clean and infected states – a hypervisor always maintains a clean uninfected image.

To contain APT attacks successfully, an out of the box approach is required. Using an embedded secure hypervisor as a proactive detection layer provides that facility.

Fig 2: How advanced persistent threats should be detected



Author profile

Avishai Ziv is vice president of cyber security solutions for Lynx Software Technologies (formerly known as LynuxWorks).