Monitoring servers using comparison of calculated versus observed power signatures

Ezekiel Kruglick, Ph.D.

***Abstract*—This paper was originally submitted to Xinova as a response to a Request for Invention (RFI) on cyber-attack forecasting, detection, and stabilization methods. In this paper, what is proposed is a security method of monitoring networking and computing devices. It describes a way of monitoring these devices by comparing the calculated versus the observed power signatures.**

I. Introduction

DETECTING a modern advanced persistent threat can be very difficult. Some extremely low-level attacks can leave behind payloads that operate below (lower security ring) even the operating system or hypervisor. Modern attacks can hide in RAM, run on supervisory processors, or run at higher level of access, for example as a hypervisor. Such ring zero (or even lower!) attacks cannot be detected or characterized by conventional tools run in the operating environment. It would be very useful to have an inescapable way to detect adversarial activity.

With all security measures, however, we desire to minimize the impact on the monitored device. Ideally, the countermeasures would also be undetectable to the attacker such that even having part-time analysis coverage would force attackers to vacate the platform or be detected.

Currently there are not many (if any) cybersecurity options that offer the combination of inescapable detection, minimal processing impact, and undetectability to the attacker. The concept presented here combines all of these advantages.

The solution presents a security system that monitors the detailed time-domain power usage of a server and compares it against known power profiles of known tasks assigned to the server to identify unaccounted resource usage and thus detect advanced intruders. In one example, a supervisory process in the datacenter may maintain a table of active tasks and loads (e.g. network traffic loads) to each server and generate predicted power profiles – the system then compares the predicted power profile to measured power usage across the servers to detect illicit resource usage that can be as slight as minor processor usage or unexpected storage access.

II. Summary of the Invention

The focus is to use direct power measurement (power channel) to gain a power signature of each device. It is interesting to note that portable devices like phone and tablets already have current and power monitoring analog chips built in to manage the battery. Supervisory information on the tasks expected or assigned to a given server or device can be quite valuable. For example, a server assigned three tasks known to the supervisor would have a time-domain power profile expectation that is a composite of the three tasks. The supervisor uses the composed time-domain expectations to compare with measured power usage to detect attackers/compromise.

III. Detailed Explanation

Analyzing power usage to detect what is happening within an electronic device has traditionally been a tool for compromising platforms, a form of attack instead of defense. Hackers have shown that power channel monitoring can identify what devices and circuits are functioning and even what operation is being performed within device logic. This is called a “side channel” attack because the attack happens outside the normal data or signal channels.

Fig. 1. Diagram from Gamaarachchi et al showing power channel attack work being performed on a smart card reader. Researchers showed an ability to extract keys this way by measuring the detailed power traces.

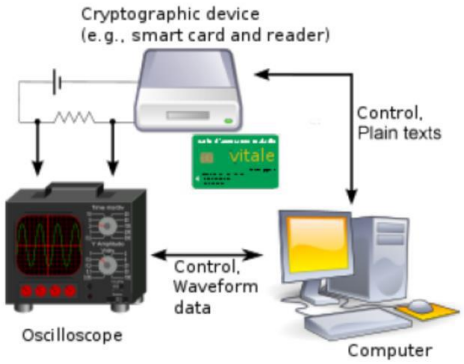
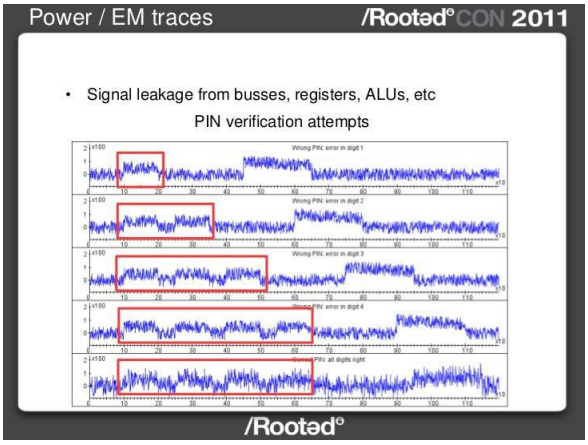


Fig. 2. A presentation from the "/RootedCon" conference showing power channel attack used to tell which pin digits are correct so that getting the whole PIN only requires a few guesses per digit instead of guessing all combinations.



None of this previous work is very useful for defense, as it all teaches ad-hoc "figuring out" what is going on within a non-compromised device. There is one security company that offers security for small IOT devices by characterizing the electromagnetic interference (EMI) it generates normally and just seeing if that typical behavior changes[1]. Such "has it changed" analysis is not useful in datacenter operations because every server in a datacenter will have a continuously changing mix of load, tasks, and programs assigned to it. Some other people have demonstrated detection of cryptographic coin mining on datacenter servers but this is based on massive overall power usage increases[2], not details of what the server is doing.

The solution here builds upon those concepts by borrowing side channel attack concepts to gather data, but using it alongside the knowledge and command structure of the datacenter to measure server power usage versus a known expected power usage profile based on assigned tasks.

On important fact this solution relies on is that a datacenter or server room provides machine operators with physical access to the devices. This allows us to install direct power monitoring. For example, a simple voltage measurement or inductive current sensor around the main power input for each rack slot may be hooked to an analog to digital convertor and then directly to serial digital data output and can be installed for under $10. Such power monitoring can detect memory access, accelerator versus CPU usage, data storage access, and many other fine details of operation.

The second interesting detail we use here is that, contrary to previous work, the datacenter or enterprise operator knows what tasks are assigned to each server. Power profiles can be generated for each task at the application level or even at the transaction level. An example of application level power profiling would be generation of profiles for each application VM that might be assigned to a server. An example of transaction level power profiling could involve monitoring incoming network traffic levels and characterizing the power usage of a particular application as a function of traffic. Transaction level power profiling provides finer detail but requires more measurement and modeling.

Finally, we maintain a supervisory security database of which collections of virtual machines or tasks are assigned to each server and build a composite expectation of power profile specific to actual assigned tasks and usage load for a particular piece of power-monitored hardware. Comparison of the actual power usage to this power expectation will reveal "hidden" processing, storage access, network access, or other resource usage. Since even hypervisor-level advanced persistent threats and firmware hacks need to use actual physical resources, such power-level monitoring can detect attacks that are otherwise currently completely undetectable.

Note that some datacenters already offer "security as a service", so there is precedent for charging people a continuing subscription cost to monitor their servers for them3.
One implementation variation might include building a suspicion that a particular machine or virtual machine is compromised and changing loads in order to more clearly evaluate the power profile. For example a machine with four virtual machines on it may generate a suspicious power signature and our system may reply by migrating the processes one at a time to other monitored hardware to see whether the suspicious signature is attached to one of the virtual machines or the actual hardware.

In another example some subset of datacenter hardware may have the monitoring hardware described in this solution, and virtual machines or applications that merit extra security (either by paying for it or due to suspicious behavior) may be purposely migrated onto the power monitored hardware for evaluation. This may be continuous evaluation or a procession of user virtual machines/tasks may be migrated through the monitored hardware for evaluation.

IV. Conclusion

While the concept is simplistic and primitive in nature, detecting consumption (CPU, memory, etc.) can be effective since these are attributes that at the moment can't be spoofed. Attackers can't perform any functions without using power. Dependencies are few, since hardware exists to measure power data on many systems already. The road to implementation could be relatively short. Proof of concept would include several steps:

1) Build power monitoring into an ordinary machine and demonstrate gathering data
2) Gather power data while running different tasks and demonstrate ability to identify changes in task load from power data.

3) Demonstrate ability to detect difference between a task running on bare metal and the same task in a virtual machine (as demonstration of detecting a hypervisor level persistent threat).

ACKNOWLEDGMENT

This research was originally submitted to Xinova, LLC by the author in response to a Request for Invention. It is among several submissions that Xinova has chosen to make available to the wider community. The author wishes to thank Xinova, LLC for their funding support of this research.

More information about Xinova, LLC is available at www.xinova.com.

REFERENCES

1. https://www.pfpcyber.com/
2.http://davidgamez.eu/papers/JinBighamRodawayGamezPhillips06_AnomalyDetectionElectricity.pdf
3.https://aws.amazon.com/blogs/aws/amazon-guardduty-continuous-security-monitoring-threat-detection/