sciendo International Conference KNOWLEDGE-BASED ORGANIZATION
Vol. XXV No 3 2019

VIRTUALIZATION SOLUTIONS SUPPORTING PRIVACY AND DATA PROTECTION IN ONLINE ACTIVITIES

Andrei ȘANDOR

NATO HUMINT Centre of Excellence, Oradea, Romania
andrei.sandor@natohcoe.org

Abstract: *Nowadays, smart devices like computers, tablets, and smartphones allow transmitting the information everywhere, with high speed, over the World Wide Web. However, risks regarding data integrity, privacy and security when using the Internet, increased dramatically, as methods designed to exploit the system's vulnerabilities are more and more sophisticated. Therefore the need for people working in professional environments to protect their private data when using unsecure connections, by employing advanced tools. There are multiple solutions, but we will focus on the use of virtualization software like VMware or Oracle Virtual Box, together with traditional privacy measures (use of proxies and VPN's). Today's smart devices store an important amount of data about their owners and, in most of the cases, people don't even realize this. Installing and using protection means is often not enough. They have to be properly setup in order to ensure the desired level of security, or anonymity, when using the Internet, and require for the military personnel a good knowledge not only about cyber vulnerabilities and risks, but also technical capabilities and features of the employed security solutions.*

DISCLAIMER*: This paper expresses the views, interpretations, and independent position of the authors. It should not be regarded as an official document, nor expressing formal opinions or policies, of NATO or the HUMINT Centre of Excellence (HCOE).*

Keywords: privacy, data protection, virtualization, virtual machine

1. Introduction

The communication and information technology is a very complex domain and offers many opportunities, but equally involves many risks. Data transmitted through the Internet can be intercepted, altered or used to track peoples' activities and interests. Nowadays, in the context of emerging cyber-threats, there are more and more discussions about privacy and data protection.

Probably one of the biggest problems regarding smart devices used by personnel working in sensitive domains is the mix of personal-related data with work-related data. For this reason, they should know how to protect their data and privacy when surfing the web.

Briefly, privacy is about how to work with a smart device without sharing data about the user of that device, regardless the actor or the reason behind.

We cannot talk about full anonymity in the Internet, but there are some technical measures and good practices that can be used in order to protect our privacy and make harder for someone to discover data about us.

The next chapter of the paper will portray aspects of privacy achieved by using the traditional technical measures and good practice regarding the use of smart devices,

DOI: 10.2478/kbo-2019-0133

while the following chapters will describe advantages of using hardware virtualization for achieving data protection. In the final part of the paper, I will emphasize the importance of privacy and data protection of the devices owned by people working in sensitive domains, where information matters.

2. Traditional privacy protection

Privacy and data protection start with basic prevention measures and computer practices like using security solutions (firewall, IDS – Intrusion Detection System), using certified software and permanent software patch updates, browsing just on websites with a trusted security certificate, using good password management, be aware of suspicious activities (phishing, social engineering, a.s.o.).

All these are basic good practices that would help us in protecting our devices from unauthorized access or data loss.

Regarding privacy, if in the past it referred mostly to hiding traces on the Internet (especially in hacking or other criminal activities), nowadays it is referred to as "data protection by design" [1]. This concept means that all data of individuals that are using information systems should be protected by legal and technical measures by default. However, currently, this concept is not fully applicable, so everyone should be aware of a minimum set of data protection measures (retrieved from common best practices) when using an internet-connected device, configuring an account on social media and using VPN's (Virtual Private Network), proxies or TOR (The Onion Router) browser for masking the IP address.

Social media has a very big influence on our lives and most of the people choose to share data on these networks. We need to be aware that all the data we upload on social media networks will be available on a server that we cannot control. An important note would be that the user can no longer totally control the access to a file (document, photo, a.s.o), once he uploaded it on a social media network, or Internet server. Even if "deleted" by the user, the file would still remain stored on the servers. So, we shouldn't upload sensitive data and also restrict other user's access to our reachable data by using the correct settings (for example, for Facebook: set up that photo are visible just for friends). However, we should be aware that national legislation may consider the private space from social platforms as public space [2] [3].

Another basic rule regarding privacy refers to the connection between the user's device and Internet. Devices should be configured to not automatically connect to available Wi-Fi; any public Wi-Fi network, whether it is free or password protected, should be treated as unsecure.

In order to safely use public networks to access Internet, it is recommended to use VPNs, proxies, the TOR network or a combination of these.

A VPN is a technology that creates an encrypted connection between two systems. This means that the activity from a system can be monitored at the edges of the encrypted tunnel: on the system or on the server that hosts the VPN (or the VPN provider).

A proxy server will forward traffic from a system to the desired destination. Proxies can hide the identity but can also be used by third party entities to log all traffic that is passed by, so not all available proxy servers can be used for privacy reasons.

TOR is a multi-proxy network that doesn't rely on specific proxy servers to process data, but uses the connections of multiple other TOR users to mask the IP of the original user. TOR is released as an open source free software, that is completely legal to use and possible to download from the TOR website.

Thus, TOR protects users' identity through a routing algorithm. So, the traffic from a system will be forwarded over multiple nodes before reaching destination. The

vulnerability of TOR is generated by compromised nodes (controlled by governmental or other entities) that make possible the discovery of an individual's activity by traffic correlation.
Each of the three measures for data protection on the Internet has strengths and weaknesses. The best solution that an Internet user can embrace to be anonymous is a combination of VPN, proxies and TOR. This can be a good solution if the slow connexion speed doesn't have a great impact on user's activity.

3. Hardware virtualization

Hardware virtualization is not a new concept and it is widely used in corporations because it reduces IT staff's workload, offers a better uptime, allows faster deployment of resources, provides energy savings and greatly improves disaster recovery.
Virtualization is a technology that allows the separation of the logical architecture from physical resources, as shown in figure 1. By controlling access to the physical resources, virtualization can be used to run different services in parallel, on the same physical hardware [4].

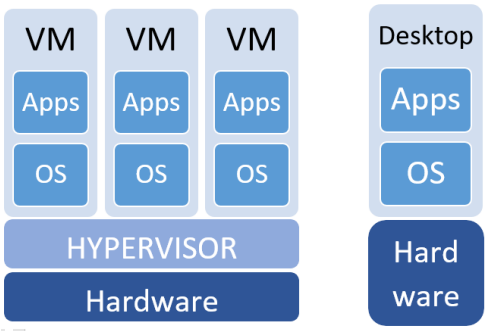

Figure 1: Virtualization vs. normal computing

The main advantages that hardware virtualization can offer to support privacy in online activities are creation/restoring of a snapshot and separation between VM (Virtual Machine) and guest OS (Operating System).
A snapshot of a VM is a saved state (resources, installed software, saved data on the virtual disk, network properties and other) of a VM at a certain time. Usually, VM snapshots are created in software development (before performing OS patching of software upgrades), but it can be used as well for privacy and data protection.
When restoring a VM to a saved snapshot, all data from the VM is lost, as shown in figure 2. This is a good point because all browsing data, cookies and potential malware or spyware are deleted and the VM is brought back at the state when the snapshot had been made. In certain special purpose configurations, you may want to exclude one or more of the virtual machine's disks from the snapshot [5]. Therefore, the virtual disk can be configured to be persistent (changes are written to disk and remain after restoring a snapshot) or non-persistent (changes are discarded when powering off the VM or restoring a snapshot).

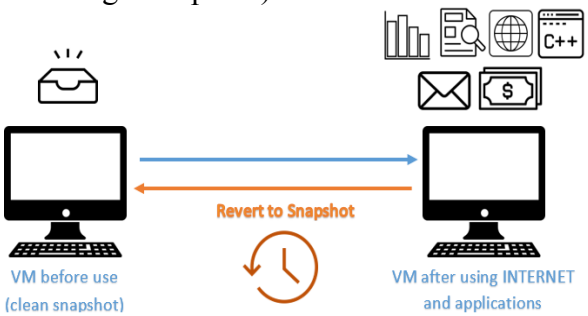

Figure 2: Restoring VM to snapshot

For privacy reasons, it is good to have an initial snapshot without browsing data. This means that after creating a VM and installing VM OS and other software, it should be created a snapshot, before starting to search the web.
Creating/restoring VM snapshots can be very useful when browsing on unsecure websites (related to hacking, for example), where the risk to access a link containing a malicious file is very high. Also, there is a risk that entities from the Internet may learn about the searches that are coming from a certain computer or from a network through spyware. With hardware virtualization, this risk still remains, but is lower if the individual working with the VMs revert

them to a snapshot based on a time schedule (for example twice a week) and use a non-persistent virtual disk.
Another important feature of hardware virtualization is represented by the separation between VM and host. This can be a total separation or a partial separation. In case of a total separation, all data inside the VM is separated from host and cannot be accessed, while in case of a partial separation, data from a VM can be accessed through an enabled USB controller, shared folders between VM and host or through a network – if the VM is configured to have attached a network adapter.
In the civilian companies, there is a trend to use personal devices for tasks related to work, known as BYOD (Bring Your Own Device). Even if the military and other public institutions have clear regulations regarding the use of information devices, sometimes sensitive data can be found on the personal devices of people working in those institutions.
The biggest problem concerning BYOD paradigm is that a device can hold both personal data and work-related data. This is a risk that can be reduced with hardware virtualization. So, for example, company servers can host VM's and employees' devices operate like remote controls when interacting with application contained on hosting hardware, and communicate via VPN connection [6]. This architecture provides a secure working environment that is accessed from potential unsecure devices. Mainly, hardware virtualization is about desktop virtualization, but mobile virtualization becomes more and more used in the communication and information sphere.

4. Mobile virtualization

Smartphones have a very important place in our lives, either if used for social media, entertainment or communication. A smartphone may know more about the owner than a computer, so data protection in these devices should be a big concern. As for computers, virtualization can be a solution for data protection and privacy in mobile devices like smartphones.
Mobile virtualization is hardware virtualization on a smartphone, enabling secure separation between the underlying hardware and the software that runs on top of it [7]. Mobile virtualization is based on the same principles as hardware virtualization, with the remarks that the VM's operating system is represented by a mobile platform (like Android).
The host is a very important element in mobile virtualization. It can be a smartphone as told before, but also a server that is storing virtual smartphones, as shown in figure 3.

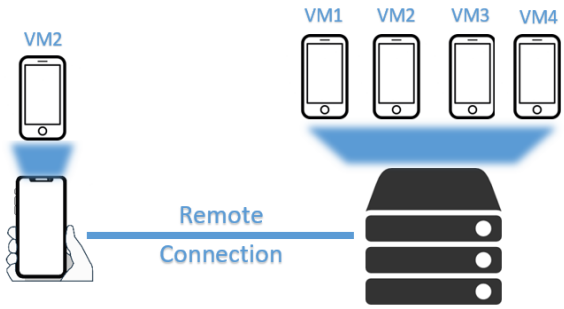

Figure 3: Remote mobile virtualization

In the first situation, with a mobile device (tablet or smartphone) as host, there are two isolated instances running their own OS environment, independently one from the other. This means that, for example, data from a mobile device can be separated so that one instance contains personal-related data and another instance contains work-related data.
Another reason for which the separation between instances is good for data protection is that if one instance is infected with malware, not all data from the mobile device is compromised.
The second situation, with a server as host for the virtual smartphones, solves many of the problems of secure containers and "bring your own device" by treating a user's physical phone as a terminal for remotely accessing a virtual smartphone running off-the-shelf smartphone apps [8].

This may keep off sensitive data in case of lost or stolen phone.

5. Conclusions

People working in sensitive domains like military or intelligence should pay a great attention to their personal data and privacy, starting from two points of view: technical data protection against cyber-threats and privacy on the Internet.

Today's cyber threats may conduct to data loss or, even worse, to data theft. Often, malware is transmitted through untrusted applications that for sure are not used and accepted by IT professionals working in public institutions, so the separation between VM's or between a VM and the host is a feature of virtualization that may come in hand to protect data. Using virtualization, people can have one personal VM for social media, browsing, personal communication, applications used for personal purpose, and another VM used for job-related issues, with accepted applications and storing just relevant data.

Big data analytics are used in various domains (advertising, economy, a.s.o.) and can uncover significant data about an Internet user starting from insignificant data. This is why people shouldn't mix personal information with work related data. At just one click distance, one can compromise his colleagues/ professional network and offer sensitive data regarding their activity, organization's standing procedures, or the institutional structure.

People need mobile and secure devices to communicate from different places. This requirement raises some challenges regarding the physical protection of the device, the security of the software environment and the security of the communication channel. Virtualization can be a solution for this requirement, solving the physical protection of the device and the security of the software environment.

Virtualization can increase the physical protection against unauthorized access because data may be stored on a server, and the host acts just like a remote device to access a VM. So, even if the host is searched for data with forensics tools, the only result will be the discovery of connection-related data.

Virtualization can solve also the security of the software environment by controlling and monitoring the virtual smartphones or desktops from a server and creating/ restoring snapshots periodically.

In order to have a secure communication channel between a remote device and a VM hosted on a server, it should be used a VPN or a series of trusted proxy servers, or both solutions, if the data link bandwidth is not the most important requirement for a specific activity.

Therefore, enhancing physical protection and maintaining a secure environment, virtualization, along with using a secure communication channel, offers people the solution they need. So, individuals may use personal or any unsecure devices to remotely open a VM hosted on a secure server and transmit data or access information.

Although technical privacy and data protection measures are very important, probably more important is the awareness about cyber vulnerabilities, risks and threats. This knowledge should be applied and translated in best practices when using smart devices and surfing the web.

References

[1] https://ccdcoe.org/library/publications/open-source-intelligence-open-social-intelligence-and-privacy-by-design-a-meta-level-rule-of-law/.

[2] https://www.hotnews.ro/stiri-esential-18726335-pagina-facebook-este-spatiu-public-nu-privat-arata-decizie-irevocabila-inaltei-curti-justitie-casatie.htm.

[3] https://uklabourlawblog.com/2018/06/22/i-lost-my-job-over-a-facebook-post-was-that-fair/.

[4] Gábor Pék, Levente Buttyán, Boldizsár Bencsáth, *A Survey of Security Issues in Hardware Virtualization,* ACM Computing Surveys, Vol. 45, No. 3, Article 40, Publication date: June 2013, p. 40:1.
[5] https://www.vmware.com/support/ws4/doc/preserve_snapshot_ws.html.
[6] Kathleen Downer, Maumita Bhattacharya, *BYOD Security: A New Business Challenge*, The 5th International Symposium on Cloud and Service Computing (SC2 2015), IEEE CS Press.
[7] Teemu Väisänen, Alexandria Farar, Nikolaos Pissanidis, Christian Braccini, Bemhards Blumbergs, Enrique Diez, *Defending mobile devices for high level officials and decision-makers*, NATO Cooperative Cyber Defence Centre of Excellence, Tallin, 2015, p. 74.
[8] https://svmp.github.io/.