An holistic approach to microsegmentation: A practical evaluation of microsegmentation in a cloud environment

Nil Ballart Cabanas

1 de juliol de 2024

Resum– Aquest projecte es basa en l'anàlisi i implementació d'una tecnologia de microsegmentació en una infraestructura cloud. La microsegmentació permet una major granularitat en les polítiques de segmentació, millorant la resiliència de l'entorn quan l'atacant ha aconseguit penetrar les defenses perimetrals i ha infectat una màquina interna. Seguint aquesta estratègia de seguretat Zero Trust reduïm en gran mesura la superfície d'atac oferint un alt nivell de control a temps real. En aquest projecte es fa un anàlisi general de les capacitats d'aquesta tecnologia, el seu impacte en els dispositius i els avantatges davant un entorn amb únicament segmentació perimetral.

Paraules clau– Microsegmentació, Segmentació, Guardicore, Zero Trust, AWS, Resiliència, Ciberseguretat

Abstract– This project is based on the analysis and implementation of a microsegmentation technology in a cloud infrastructure. Microsegmentation allows for greater granularity in segmentation policies, improving the resilience of the environment when the attacker has managed to penetrate the perimeter defenses and has infected an internal machine. Following this Zero Trust security strategy, we greatly reduce the attack surface by offering a high level of real-time control. This project provides a general analysis of the capabilities of this technology, its impact on devices, and the advantages over an environment with only perimetral segmentation.

Keywords– Microsegmentation, Segmentation, Guardicore, Zero Trust, AWS, Resilience, Cybersecurity

1 Introducció

La protecció d'infraestructures[1] es basa en securitzar infraestructures d'aplicatius de negoci i protegir tecnologies, sistemes i actius d'una empresa. Els sistemes engloben des de xarxes sense cable per connectar-se a Internet fins a solucions cloud, servidors, dispositius mòbils, etcètera.

La seguretat de les infraestructures sempre ha estat un dels pilars més importants en la ciberseguretat. Aquesta àrea es troba dintre del anomenat "*Blue Team*". La protecció d'infraestructures té com a objectiu principal protegir la xarxa interna d'una empresa i els seus actius davant atacs, accessos no permesos i altres accions malicioses enfocades en perjudicar l'organització. Avui en dia, hi ha un gran ventall de tecnologies molt eficaces i en constant millora que permeten aixecar un mur de protecció davant atacs provenint de l'exterior de l'empresa o organització.

Encara que aquestes tecnologies de protecció perimetral tinguin una eficàcia molt alta, cap protecció és perfecte i sempre hi haurà vulnerabilitats, descobertes o no, que permetran a un atacant realitzar accions malicioses. És per això que en els últims anys, sobretot després de la gran transformació digital que hi ha en curs i, juntament amb la transició a una infraestructura híbrida en cloud i l'alta proliferació de ransomwares, s'ha començat a buscar solucions complementàries a aquesta defensa perimetral, que protegeixin a la mateixa escala la xarxa interna davant aquests atacs, tot amb un impacte mínim en el rendiment dels sistemes i amb adaptabilitat als canvis.

1.1 Estat de l'art

Actualment, aquesta protecció afegida la podem aconseguir mitjançant un model de seguretat Zero Trust [2]. Aquest model és relativament complex i costós d'aplicar i mantenir, essent aquestes dos les principals raons per les quals les

• E-mail de contacte: nil.ballart.c@gmail.com
• Mención realitzada: Tecnologies de la Informació
• Treball tutoritzat per: Marc Ortega Gil (Departament d'Enginyeria de la Informació i de les Comunicacions)
• Curs 2023/24

empreses no tenen microsegmentació [3].

1.1.1 Zero Trust

El model de seguretat Zero Trust ja porta un temps donant-se a conèixer, apareixent per primer cop al 2010 [4]. Aquest es basa en canviar la forma de pensar alhora de protegir les infraestructures. Passem d'una estratègia de çonfiar, però verificar"a "no confiar mai i comprovar sempre". Això implica que no es confia amb cap usuari ni dispositiu per accedir a un recurs fins que es comprovi la seva identitat i s'autoritzi. Aquest procés s'aplica a totes les connexions, incloent les que tenen origen en la xarxa interna. Per dur a terme aquesta estratègia, hi han diverses maneres d'aconseguir els resultats esperats, però la principal manera que existeix actualment i les més complerta és mitjançant una solució anomenada Secure Access Service Edge (SASE) [5].

1.1.2 Secure Access Service Edge (SASE)

SASE va sorgir com un paradigma de seguretat en el núvol que va més enllà de les xarxes tradicionals, oferint una integració de funcionalitats per protegir l'accés a aplicacions i dades de forma segura, seguint aquest model de Zero Trust. Una solució és considerada SASE quan integra en aquesta les següents capacitats, essent només les dues primeres necessàries i les posteriors opcionals:

- **Secure Web Gateway o proxy web (SWG)**: Aquesta solució actua com a protecció entre la xarxa interna i L'Internet públic, filtrant i controlant el tràfic web per prevenir amenaces com el malware, phishing i ransomware. Al integrar-se amb SASE, es converteix en un punt d'inspecció centralitzat, oferint visibilitat i control granular sobre l'accés a llocs web i aplicacions SaaS.

- **Zero Trust Network Access (ZTNA)**: Aquesta solució redefineix l'accés a la xarxa, eliminant la tradicional confiança implícita en la xarxa interna. Això ho fa verificant la identitat i el context de cada usuari i dispositiu abans d'atorgar accés a recursos específics, minimitzant així la superfície d'atac i prevenint accessos no autoritzats.

- **Software Defined WAN (SD-WAN)**: Aquesta arquitectura de xarxa transforma la forma en que les empreses es connecten a les seves xarxes. SD-WAN separa el pla de control (que defineix com s'encamina el tràfic) del pla de dades (que transporta el tràfic real), permetent així una gestió centralitzada i flexible de la xarxa.

- **Microsegmentation:**

La microsegmentació [6] ofereix diversos avantatges davant enfocaments més establerts, com la segmentació de xarxes i d'aplicacions mitjançant components hardware (routers, switches, etc). Els mètodes tradicionals depenen en gran mesura dels controls basats en la xarxa, que són imprecisos i, sovint, difícils de gestionar. Envers això, la microsegmentació és un element de segmentació basat en software, que ofereix flexibilitat per ampliar la protecció, visibilitat dels sistemes en qualsevol lloc i, tot això, amb un alt nivell de detall i més granularitat en els controls de seguretat. Aquest complement està enfocat principalment a protegir xarxes internes on-premise, ja que normalment els entorns cloud ja tenen una microsegmentació aplicada bàsica (AWS: Security Groups, Azure: Network Security Groups, OCI: Security Lists, i altres).

1.2 Context del treball

A l'hora de la pràctica, totes aquestes millores de seguretat i avantatges necessiten d'anar acompanyades de dades que les demostrin i exemplifiquin el potencial de les seves capacitats. Aquí és on el meu treball està enfocat. Partint d'una infraestructura operativa en un entorn cloud, en aquest projecte busco desplegar i configurar un software que permeti aplicar aquesta microsegmentació en un entorn i extreure'n conclusions objectives, aprofundint en les capacitats de resiliència davant atacs basats en moviments laterals com el ransomware, ïsland hopping"[7] o l·living off the island"[8].

2 OBJECTIUS

En aquest treball busca realitzar un anàlisi teoricopràctic de les capacitats, avantatges i impacte d'una tecnologia de microsegmentació en una infraestructura de xarxa en un entorn cloud on només hi ha una segmentació plana. A través de l'anàlisi teòric de les capacitats de la microsegmentació i l'aplicació al funcionament d'una infraestructura ja desplegada i operativa, es busca obtenir unes dades objectives de les oportunitats que aquesta tecnologia pot aportar a la millora de la seguretat i la resiliència de les empreses. Per tal d'aconseguir aquest objectiu, cal assolir una sèrie d'objectius bàsics:

- Desplegar i configurar la tecnologia de microsegmentació.

- Realitzar un proves de seguretat en la infraestructura sense microsegmentació i després d'aplicar-la, comparant i analitzant els resultats.

- Desplegar un servidor Command&Control de la tecnologia i altres elements necessaris per a la seva correcta operabilitat.

- Analitzar l'entorn i les capacitats de la tecnologia per dissenyar i aplicar polítiques de microsegmentació.

- Automatitzar el procés d'anàlisi de comunicacions mitjançant un script que tracti les dades recollides de l'API de la tecnologia i generi un document Excel amb aquestes dades.

3 METODOLOGIA I PLANIFICACIÓ

Per assolir els objectius del projecte, s'ha dividit aquest en dues fases. La primera ha estat més aviat una part prèvia basada en la preparació i configuració de l'entorn cloud i l'estudi a fons del funcionament i el codi de l'eina per fer les proves de seguretat. Aquesta fase ha consistit en desplegar i configurar els elements d'infraestructura necessaris en AWS, per tal de poder instal·lar i desplegar posteriorment els agents necessaris als sistemes sobre els que he realitzat les proves.

Per altra banda, la segona fase ha seguit una metodologia incremental. Aquesta fase ha consistit en realitzar els proves de seguretat, analitzar els resultats, aplicar polítiques i altres configuracions de microsegmentació i, finalment, tornar a realitzar les mateixes proves de seguretat per poder extreure resultats.
La idea d'aquest apartat era seguir el procés típic que es fa en projectes de definició de polítiques, que normalment consisteix en 4 fases:

1. Definir quines comunicacions permetre, alertar i bloquejar segons les necessitats i objectius de seguretat de la infraestructura.

2. Configurar les polítiques corresponents i, en comptes de configurar les polítiques de bloqueig directament, configurar-les en mode d'alerta durant un període de prova, per tal no bloquejar erròniament alguna comunicació.

3. Analitzar les alertes de les polítiques de bloqueig que estan en procés de prova i generar les excepcions corresponents.

4. Després d'un període de prova, passar les polítiques d'alerta definides a bloqueig.

Aquest procés es repeteix fins aconseguir els objectius desitjats i cobrir totes les necessitats de seguretat. Aproximadament cada cicle incremental té una durada de 2 setmanes, on la primera és la de disseny i configuració de polítiques i la segona és el procés de prova i l'anàlisi dels resultats.

Així doncs, seguint tots els apartats definits prèviament, he anat realitzant les tasques en cicles definits d'una setmana, tal com es pot observar en el diagrama de Gantt del projecte (apèndix A1, Fig. 12).

4 Tecnologia de microsegmentació: Guardicore

4.1 Què és la microsegmentació?

Quan parlem de microsegmentació ens referim a segmentació basada en software. Aquest software s'instal·la en els dispositius i és el què permet controlar les comunicacions entrants i sortint en temps real. Al tenir un agent instal·lat en cada dispositiu, podem arribar a definir una segmentació a nivell de processos. A més a més, podem afegir diverses etiquetes a cada màquina per poder estructurar i organitzar tot el conjunt d'actius, creant una estructura lògica indiferent a la localització de cada dispositiu (física, cloud o dinàmica). Amb tot això, tenim que les polítiques de microsegmentació es poden definir incloent grups d'etiquetes, etiquetes, protocols de comunicació, ports i processos tant a nivell d'origen com de destí.

4.2 Solucions de microsegmentació líders

Actualment, la majoria de solucions de microsegmentació estan encara en desenvolupament constant, per el què trobem poques solucions completes. Per exemple Cisco Secure Workload [9] és una solució de Cisco però que no es pot

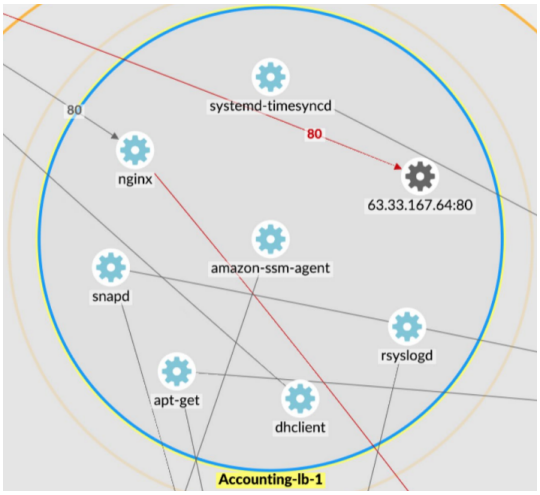

Fig. 1: *Mostra de la granularitat de manera gràfica de Guardicore Centra*

contractar individualment, sinó que ve juntament amb alguns paquets de tecnologies com un afegit, no una solució completa.

Fig. 2: *Gartner Magic Quadrant de solucions de Microsegmentació*

Les solucions líders actualment hi trobem Guardicore [10] (comprada fa uns anys per Akamai Technologies) i Illumio. La principal diferència entre aquestes dues és que Illumio [11] no requereix de la instal·lació d'un agent, juntament amb els seus avantatges i desavantatges, però això fa que siguin dos productes bastant diferents. L'altre producte que s'aproxima a Guardicore és ColorTokens [12], però aquest està enfocat a ser una microsegmentació molt simple basada en agents que puguin executar-se en qualsevol SO, establint-se com una de les solucions més adequades per fàbriques, on normalment tot el parc de dispositius té SO antics que són molt difícils d'actualitzar i/o modernitzar.

4.3 Com funciona Guardicore (Akamai Technologies)

La tecnología de Guardicore té 4 elements principals:

- **Servidor de control**
  És el servidor central que serveix per controlar i gestionar tots els dispositius amb l'agent instal·lat. Es pot desplegar de manera AIO (All in one) o com un servidor distribuït tant On-premise com al Cloud. Es pot gestionar tant a través de la CLI integrada com a través del servei web hostejat localment, el qual és el recomanat per treure'n el màxim profit.
- **Agregadors**
  Aquestes màquines es despleguen en xarxes on tant el servidor de control i les màquines que tinguin l'agent s'hi puguin comunicar. Aquestes serveixen d'intermediaris entre el servidor de control i els agents. Poden suportar fins a 500 agents alhora.
- **Agents**
  Aquest és el software que s'instal·la en cada màquina per poder controlar les comunicacions. Les màquines necessiten tenir compatibilitat de SO amb l'agent, el qual pot portar problemes amb màquines bastant antigues. Si el SO no és compatible, es pot instal·lar igualment l'agent però només en mode visibilitat, sense la capacitat d'aplicar polítiques.
- **Col·lectors**
  Aquestes màquines s'instal·len quan hi ha màquines que no poden tenir l'agent instal·lat. Aquestes fan un control de les comunicacions a nivell de xarxa, de manera semblant a la Illumio.

5 EINA DE PROVES: INFECTION MONKEY

L'Infection Monkey [13] és una eina de seguretat opensource propietat d'Akamai Technologies. Aquesta eina s'instal·la en un servidor dedicat des d'on es poden gestionar i controlar simulacions d'atacs que es poden duur a terme. També, allotja un servidor web intern on es pot accedir en el seu front-end (en l'apèndix A1, fig 13 es pot veure la seva arquitectura i comunicacions).
Les diverses simulacions d'atacs que permet l'eina es poden configurar per determinar l'abast de propagació i les màquines amb les que volem realitzar l'estudi, de tal manera que en cas d'entorns de producció o entorns crítics podem deixar fora certes màquines de manera preventiva.
Aquesta eina és divideix en dos components:

- **Monkey Island**
  Servidor dedicat per controlar i visualitzar el progrés de la simulació d'atac dintre de la xarxa. Per defecte, el servidor té un certificat self-signed que utilitza per comunicar-se de forma segura amb els agents mitjançant HTTPS.
- **Infection Monkey Agent**
  Aquest és l'agent encarregat de simular atacs en les màquines infectades i reportar al Monkey Island. Els atacs es poden iniciar en el servidor dedicat o bé descarregant l'agent en alguna màquina, simulant la infecció d'aquesta.

5.1 Procés i configuracions d'atac

El procés d'atac que realitza l'Infection Monkey [14] és simple d'entendre. Primerament intenta realitzar moviments laterals des de la màquina on es troba mitjançant un seguit d'exploits (Log4Shell i Zerologon) i atacs de força bruta en diferents processos (PowerShell, WMI, SSH, SMB, MSSQL, Hadoop, RDP i SNMP). Un cop un atac té èxit, s'executa l'atac o atacs definits en la configuració i llavors es descarrega el binari des del servidor de control, per tornar a llançar el mateix atac des de la màquina infectada.
Encara que la microsegmentació no s'encarrega de protegir què passa dintre de les màquines, només les seves comunicacions entrants i sortints, és interessant saber quins atacs es poden simular per poder fer altres proves de seguretat amb altres tecnologies. Les simulacions d'atac possibles són les següents:

- **Ransomware Simulation**
  Aquest escenari representa un atac ransomware en la xarxa. Aquest ransomware realitza accions totalment segures per entorns de producció, ja que només encripta els fitxers que es troben en una carpeta concreta que podem seleccionar prèviament. A més a més, la encriptació que fa és un bitflip i l'addició de la extensió "*.m0nk3y*", el qual és fàcilment detectable i reversible.
- **Network Breach**
  Aquest escenari simula la intrusió d'un atacant a la xarxa interna i és el que s'encarrega de realitzar els moviments laterals utilitzant vulnerabilitats conegudes, força bruta i altres exploits coneguts com Mimikatz, que permet robar credencials per utilitzar-los en moviments laterals.
- **Credentials Leak**
  Aquest escenari requereix d'introduir credencials d'algun usuari o servei i permet veure quin impacte tindria el robatori d'aquestes. A més, permet habilitar l'intent de robar credencials SSH emmagatzemats en el sistema.
- **Cryptojacker Simulation**
  Aquesta escenari simula l'execució d'un programa dedicat a minar Bitcoin que ha estat instal·lat de manera no autoritzada. Això ho fa augmentant l'ús de RAM i CPU fins a uns nivells que es poden configurar i també genera tràfic de xarxa idèntic al de Bitcoin. S'ha d'anar amb compte amb aquest escenari ja que sí que pot afectar al funcionament de màquines crítiques o de producció.

En l'apèndix A1, figura 14 es pot observar el resum d'aquest procés d'atac i els seus mètodes.

6 ANÀLISI PRÀCTIC EN AWS

Aquest apartat té com a objectiu entendre millor el funcionament i posar a prova les capacitats de la microsegmentació. Per això he desplegat un seguit de màquines en AWS que em serviran per representar un entorn híbrid on fer les proves de seguretat.
Abans d'això, exemplificaré un projecte de microsegmentació i quin és l'objectiu al que es pot arribar amb aquesta tecnología.

6.1 Entorn d'exemple

Avui en dia la majoria d'empreses han partit des d'una infraestructura purament On-premise i poc a poc van migrantse al Cloud, ja que aquest ofereix una reducció de costos, escalabilitat, flexibilitat i una millor accessibilitat. Un entorn on hi trobem infraestructura tant al Cloud com On-premise s'anomena una infraestructura híbrida, i aquesta és la que farem servir com exemple.
Imanigem una empresa de tamany mitjà amb la següent arquitectura:

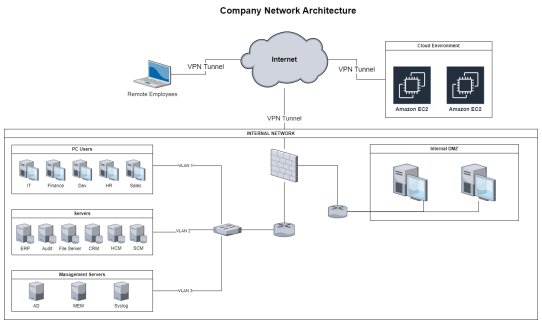

Fig. 3: *Arquitectura d'exemple*

En aquesta hi trobem els elements típics actualment. Per una banda, tenim la xarxa interna amb la seva DMZ, que conté dos serveis exposats a Internet com poden ser serveis web, i llavors diferents VLANs amb ordinadors dels treballadors, servidors interns i servidors de control i gestió, que són més crítics. Per altra banda trobem dos màquines EC2 d'AWS i un firewall configurat permeten connexions VPN per els empleats que treballin de manera remota.

6.1.1 Segmentació física

Encara que aquest entorn híbrid es podria arribar a reduir significativament la superfície d'atac dels moviments laterals amb segmentació, el descontrol que està emergent amb la migració de les infraestructures en el Cloud, les aplicacions SaaS i altres serveis en el Cloud, dificulta molt el control de tots els elements de l'organització i augmenta la superfície d'atac a les que les empreses estan exposades. A més a més, amb l'alta proliferació que hi ha hagut últimament, es veu clarament que un cop s'infecta un servidor intern, la majoria de vegades acaba amb resultats crítics, afectant a tota la organització.

6.1.2 Segmentació lògica

Seguint l'exemple anterior i el procediment típic per aplicar microsegmentació, podríem definir la següent metodologia d'etiquetatge després de desplegar agents en tots els dispositius, creant la següent segregació lògica per etiquetes:

A partir d'aquí s'han d'analitzar comunicacions i definir polítiques. Per simplificar això es poden crear grups d'etiquetes, etiquetes del estil: "Entorn: Desenvolupament", "Entorn: Producció", etc. Tot definint quina segmentació és vol configurar. Per exemple, unes polítiques de microsegmentació sobre el servidor de IT de l'exemple, podrien ser les següents:

Aquesta part d'anàlisi de comunicacions i definició de polítiques es pot fer mitjançant el mapa visual que proporciona l'eina o exportant els logs en format CSV. Si el projecte

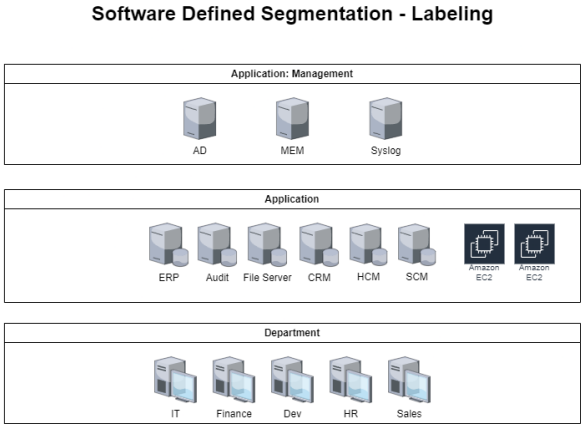

Fig. 4: *Segregació lògica*

Source	Destination	Enforcement
Department:IT Port 3389 *mstsc.exe (RDP)*	Application:AD	Allow
ANY	Application:AD	Block
Deprtment:IT 22,80,443,3389 TCP	ANY	Allow
Department:IT	ANY	Block
ANY	Department:IT	Block
ANY	ANY	Block

TAULA 1: POLÍTIQUES D'EXEMPLE

engloba molts dispositius, és molt difícil treballar sobre el mapa interactiu i el fitxer de logs és desmesuradament gran i, en aquests casos, és quan entra l'script d'automatització d'aquest procés, del qual parlo més endavant.

6.2 Exemple basat en un projecte real

La capacitat de segregar comunicacions internes de manera tant granular (fins al nivell de procés) permet grans millores en certs escenaris actuals. Per entendre una mica millor aquestes millores, a continuació explico un exemple basat en un cas real d'un projecte:
- Una companyia té desplegat l'EDR de Microsoft en tots els seus dispositius (Microsoft Defender for Endpoint), el qual és un dels més utilitzats en entorns on es treballa purament amb màquines Windows. Aquest agent s'instal·la en totes les màquines igual que Guardicore i també reporta i rep comunicacions d'un servidor intern. Aquest control es fa a través de RDP (port 3389), fent que totes les màquines hagin de tenir habilitada aquesta connexió, tal com podem observar en els requeriments a la pàgina oficial [ToDo Actualitzar referència]:

Això vol dir tenir en totes les màquines obert el port

Process	Protocol	Port	From	To
RDP	TCP	3389	Defender for Identity Sensor	All

TAULA 2: POLÍTICA D'EXEMPLE

3389 per poder rebre i enviar comunicacions d'aquest sensor legítim amb propòsits de seguretat. Aquesta exposició pot ser mitigada definint una política que permeti únicament a aquest procés la comunicació en les màquines on es té instal·lat aquest software.

6.3 Entorn desplegat

El desplegament que he realitzat es divideix en 2 parts. Per una banda tenim el servidor de control distribuït de Guardicore i, llavors, un seguit de màquines Windows Server i Linux Red Hat, en les quals llençaré els atacs i el intents de moviments laterals. Guardicore Centra, que és com s'anomena al servidor de control, està en una única VLAN d'AWS, les quals són anomenades VPC. La resta de màquines es troben repartides entre aquesta i una altre VPC. És important entendre que el servidor de Guardicore es troba juntament amb la resta de màquines, però aquest està fora de l'abast de les proves de seguretat que he realitzat.

6.3.1 Servidor de control distribuït

Aquest compta de 8 instàncies EC2, on cada màquina s'encarrega de diferents funcions de la solució.

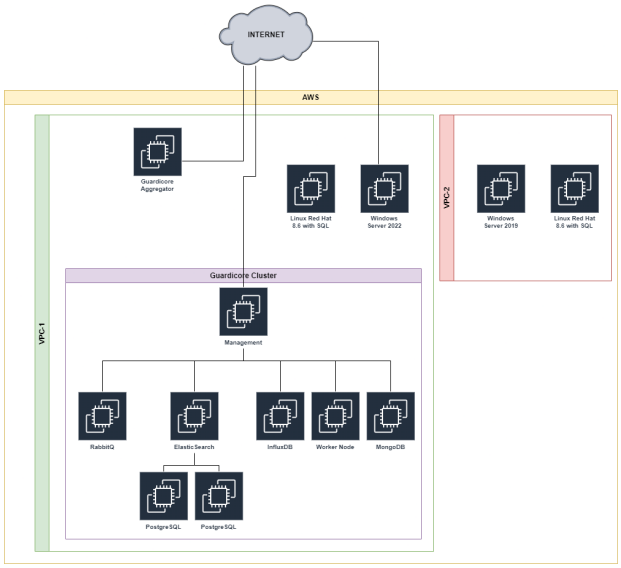

Fig. 5: *Arquitectura de les màquines desplegades en AWS*

Primerament tenim el nucli del servidor distribuït, que és el Management Control Server. Llavors, tenim 2 Workers, que bàsicament serveixen per balancejar la càrrega de treball del node principal.

La resta de màquines són les encarregades de processament de mapes, emmagatzematge de logs i dades i altres funcions que no entraré en detall. Aquestes màquines són:

- RabbitMQ
- InfluxDB
- ElasticSearch
- 2x PostgreSQL (amb 1TB d'emmagatzematge cada node)
- 2x MongoDB (configurades en alta disponibilitat)

I per últim, un agregador, cap al qual s'hauràn de comunicar i rebre comunicacions de les màquines que tinguin l'agent instal·lat. Aquesta comunicació es fa mitjançant HTTPS(443) amb SSL i un certificat autofirmat que ve per defecte, encara que aquest es pot canviar.

6.3.2 Entorn representat

Les màquines de l'entorn real que estic representant són imatges planes de diferents SO, excepte d'un Windows Server 2019 on hi tinc instal·lat el programari per dur a terme les simulacions dels atacs, els quals entraré en detall més endavant. Aquests SO són:

- Windows Server 2022
- Windows Server 2019 (Monkey Island)
- Linux Red Hat 8.6 amb SQL
- Linux Red Hat 9

Encara que aquestes màquines estiguin situades en AWS, les he configurat seguint el primer exemple d'un entorn híbrid (Fig. 2: Arquitectura d'exemple). El Windows Server 2019 representa un servidor a la DMZ exposat a internet i, per tant, a atacs de l'exterior. La resta de màquines es poden comunicar per SSH (22) o RDP (3389), HTTP/HTTPS (80, 443) i FTP (20, 21) entre elles.

A totes aquestes màquines els hi he instal·lat l'agent de Guardicore i he generat moviments laterals legítims entre elles tant per SSH com RDP.

6.4 Proves inicials de seguretat

6.4.1 Simulació d'un atac a la màquina de la DMZ

La simulació representa un atac un cop ha sobrepassat la defensa perimetral del firewall i ha aconseguit infectar una màquina a la DMZ, la qual estava exposada internet publicant un servei web o similar.
Aquest ha estat configurat perquè realitzi les següents accions:

- Network scan a les dos subxarxes (VPC-1 i VPC-2)
- Que s'intentin tots els exploits de moviments laterals, extracció de credencials i atacs de força bruta
- Que l'Infection Monkey es propagui amb un màxim de 2 salts laterals

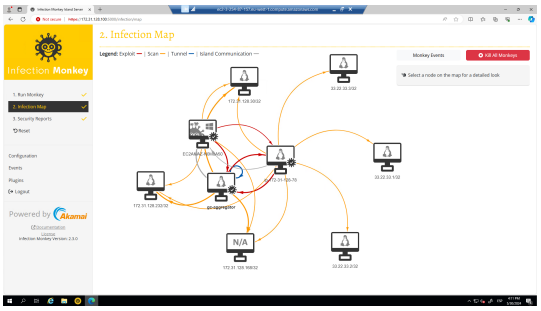

Fig. 6: *Resultat gràfic de la simulació*

Per una banda, es pot observar que es té visibilitat completa de tota la xarxa interna. A més a més, s'ha pogut vulnerar 2 comunicacions SSH (mitjançant el robatori de credencials) i una comunicació RDP.

6.5 Anàlisi de les comunicacions

Ara, un cop tenim el tràfic tant dels moviments laterals legítims com dels generats per l'Infection Monkey, podem analitzar aquest per veure què hauríem de bloquejar i què podem permetre. Aquest punt és bastant crític i, en projectes reals, s'ha de fer molta recerca i entendre realment què està passant, ja que mai serà tant senzill com el cas sobre el que estic treballant.

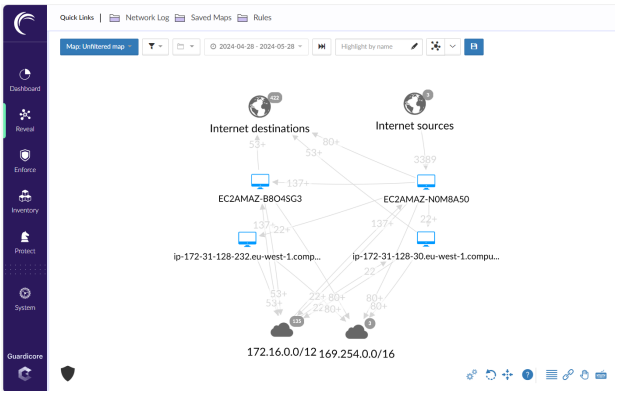

Fig. 7: *Mapa no tractat que mostra Guardicore Centra*

Principalment, com es pot observar en l'apartat anterior, els exploits exitosos de l'atac han estat a través de RDP i SSH. L'exploit de Powershell és una execució de shell remota, però només existeix un cop la màquina ha estat compromesa, així doncs, la deixarem de banda ja que ens volem enfocar en mitigar el màxim les vulnerabilitats en els moviments laterals.

Per entendre millor què ha passat i d'on prové la vulnerabilitat, mitjançant l'análisi del informe generat per l'infection monkey i les comunicacions entre les màquines, es pot observar com, primerament, aconsegueix robar les credencials SSH del Windows Server 2019 (el que es troba a la DMZ i des d'on s'ha iniciat l'atac). Llavors, utilitza aquestes credencials per explotar per força bruta tant les comunicacions RDP com SSH.

Un cop analitzat què ha passat, tenint aquestes comunicacions indesitjades i les que volem permetre per tal de que la nostra empresa fictícia pugui seguir en funcionament, hem de configurar polítiques que només permetin el que nosaltres considerem com a legítim.

6.6 Configuració de les polítiques

En aquest apartat és on es pot observar realment les capacitats de microsegmentació envers la segmentació. Sense microsegmentació, un moviment com hem pogut observar en l'atac simulat a través de SSH o RDP des d'una màquina a una altra només el podríem bloquejar o permetre si les màquines es trobessin en diferents VLAN i la infraestructura on estiguessin desplegades ho permetés. Encara que poguéssim realitzar aquest control, només tenim la opció la bloquejar o permetre la totalitat de les comunicacions, sense anar més enllà.

En canvi, amb microsegmentació podem aplicar una configuració molt més granular, definint origen i destí fins a nivell de procés. Això ens serveix per definir, per exemple, unes polítiques que bloquegin les comunicacions SSH a no ser que es realitzin mitjançant l'aplicació de PuTTy (o qualsevol altre que realitzi aquesta funció). A més a més, podem definir si aquest programa es pot trobar instal·lat en qualsevol directori o en el directori per defecte que es troba el programa o procés.
Per tant, per poder permetre els moviments legítims i bloquejar els maliciosos que, en la majoria de casos, al ser automatitzats, no s'executaran fent servir aquests programari que hem mencionat.
Aquí podem veure les polítiques configurades, on tenim per una banda bloquejos generals de les comunicacions i llavors les comunicacions explícitament permeses que he comentat.

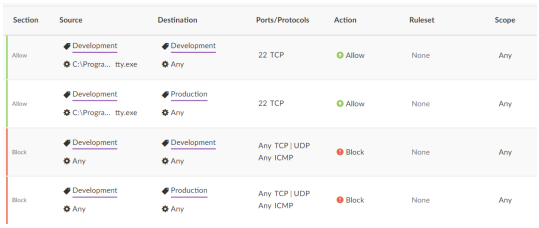

Fig. 8: *Polítiques permeten només comunicacions SSH a través de PuTTy*

A part, també he bloquejat el protocol ICMP, per tal d'evitar que es tingui visibilitat de les màquines que no desitgem des de dins de la xarxa.

Un cop activades aquestes polítiques, he verificat que les comunicacions mitjançant SSH per consola no es poden realitzar i les connexions a través de PuTTy sí.

Hem de tenir en compte que en projectes reals aquestes polítiques no s'haurien de crear directament en mode bloqueig, sino que s'haurien de crear en mode alerta per poder analitzar les comunicacions i generar les polítiques de la forma més granular possible.

6.7 Proves finals de seguretat

6.7.1 Simulació d'un atac a la màquina de la DMZ

Un cop aplicada la microsegmentació en l'entorn podem tornar a simular què passaria si una de les màquines exposades a internet que es troba a la DMZ és infectada.
Primerament, observem com apareixen notificacions de que algunes comunicacions estan essent bloquejades.

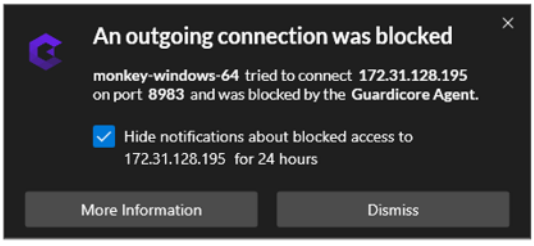

Fig. 9: *Missatge que ens adverteix que hi ha comunicacions sortints que s'estan bloquejant*

Un cop acabat l'atac, podem observar un resultat totalment diferent.

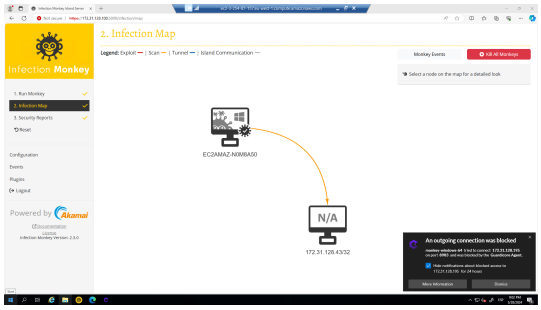

Fig. 10: *Resultat gràfic de l'atac un cop aplicada la microsegmentació*

Podem observar com el network scan executat per l'Infection Monkey aquesta vegada ha detectat únicament l'agregador de Guardicore, però sense poder-hi accedir com en l'últim atac. La resta de màquines que s'havien detectat en l'atac inicial ni apareixen com a possibles objectius de moviments laterals ja que aquests no són detectats, almenys per l'Infection Monkey.

Com que la idea dels atacs és poder demostrar com la microsegmentació és capaç de reduït significantment els atacs laterals, he deshabilitat el bloqueig del protocol ICMP per veure si, encara que el malware que hagi infectat a la màquina de la DMZ detecti totes les màquines de la xarxa interna, aquest no és capaç de vulnerar-les com havia fet anteriorment.

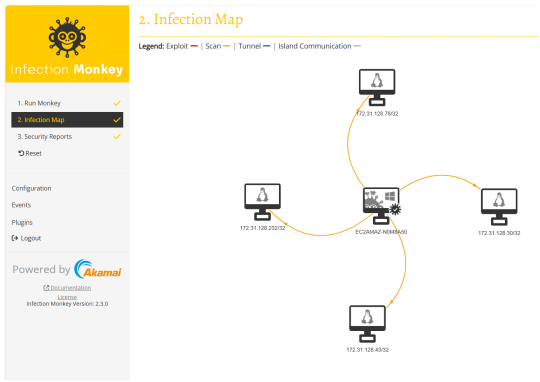

Fig. 11: *Resultat gràfic de l'atac un cop aplicada la microsegmentació i desactivat el bloqueig ICMP*

7 SCRIPT D'AUTOMATITZACIÓ DE L'ANÀLISI I TRACTAMENT DE DADES

Actualment, un dels apartats més manuals que hi ha al realitzar un projecte de microsegmentació és l'anàlisi de les comunicacions i la definició de quines polítiques es volen crear. Més enllà de haver d'entendre el funcionament de totes les màquines d'una organització (servidors, portàtils, impressores, ordinadors de sobre taula, thin clients, dispositius IoT, etc.) també cal entendre cada una de les comunicacions entrants i sortints de les màquines que es volen protegir amb microsegmentació.

Per fer-ho, de moment, hi ha només dues opcions. O bé es pot fer utilitzant el mapa interactiu, el qual et permet seleccionar una comunicació i crear una regla a partir d'una comunicació o bé descarregant els logs i, mitjançant excel o alguna eina per tractar dades CSV, fer un anàlisi més exhaustiu.

És cert que la opció del mapa interactiu és molt senzilla i és la que més es sol utilitzar, però quan tenim entorns amb una gran quantitat de dispositius (més de 500) aquest mapa i totes les comunicacions, que es representen amb fletxes, és molt difícil de manejar. És per això que en aquests casos s'opta per la segona opció. Aquí és on actualment, almenys en els projectes que tinc visibilitat, s'ha de treballar amb aquestes dades CSV (més de 200.000 línies) de totes les comunicacions de certes màquines durant 30 dies, el qual és un procés bastant tediós i feixuc que pot arribar a durar setmanes.

Per automatitzar i agilitzar aquest procés, he creat un petit script que es comunica amb l'API de Guardicore. Un cop autenticat amb un usuari de servei de només lectura que he creat per aquest propòsit, faig unes queries i permet-ho triar a l'usuari sobre quina etiqueta obtenir les comunicacions. Un cop això, demano mitjançant l'API totes les comunicacions entrants i sortints d'aquesta etiqueta en el mapa que s'ha creat amb la informació de l'interval de temps que s'hagi esperat (normalment 30 dies) i faig un tractament de dades senzill per tenir una base molt més comprensible i compacte per poder entendre les comunicacions i definir les polítiques que es volen aplicar.

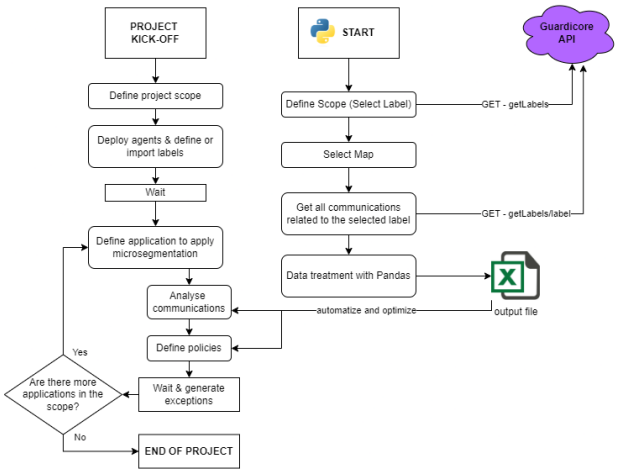

Fig. 12: *Diagrama del funcionament del programa desenvolupat i en quin apartat del cicle d'un projecte entra en funcionament*

8 ALTRES CAPACITATS

8.1 Mapa de comunicacions en temps real

Com ja s'ha pogut observar en les fotos dels mapes de comunicacions generats per Guardicore, aquesta eina, a part de aplicar microsegmentació, ens permet tenir un mapa en temps real de totes les comunicacions que s'estan realitzant. Aquesta característica per si sola aporta molt valor ja que normalment les empreses no tenen aquesta visibilitat. Amb

això, podem detectar fàcilment problemes de seguretat existents i potencials que no tenen perquè tenir a veure amb el què estem mitigant amb la microsegmentació.

8.2 Isolació dinàmica de sistemes

Una altra capacitat que té la tecnologia és, aprofitant que tenim un agent instal·lat revisant totes les comunicacions, monitoritzar constantment els processos interns per intentar trobar, mitjançant IA, anomalies i potencials malware. Un cop algun comportament d'aquests és detectat, s'envia l'execució d'aquest procés a una màquina virtualitzada perquè actui com a honeypot alhora que analitza el comportament. Això pot arribar a evitar 0-days i ransomwares.

8.3 Segmentació senzilla en xarxes planes

És molt comú trobar-se en l'àmbit de la ciberseguretat empreses petites i mitjanes que, per una o altra raó, tenen una xarxa interna plana, essent aquest un dels seus principals punts crítics davant d'atacs. Normalment, un reorganització de l'arquitectura interna és costosa i complexa, però si es vol tenir una bona seguretat i mitigar l'impacte d'una intrusió a la xarxa interna això és el primer que s'ha de fer.

Això pot tenir una ràpida solució aplicant una segmentació bàsica, sense arribar a nivell de processos ni ports. Aquesta segmentació basada en software que ens ofereix Guardicore pot solucionar molt més senzillament aquest problema i, a més a més, deixar marge de millora per arribar a aplicar microsegmentació en un futur.

8.4 Segregació lògica per auditories

Faciliten l'auditoria de comunicacions de les màquines dintre cert abast.

9 Conclusions

La microsegmentació és poc coneguda i relativament nova, però és una solució que cada vegada té més potencial davant els problemes de seguretat actuals.
Entendre bé què és la microsegmentació i quins beneficis de seguretat té, pot ajudar molt a aportar valor i millorar la seguretat de les empreses. És cert que aquesta no ha de ser una solució única ni molt menys, ja que sempre ha d'anar acompanyada d'altres proteccions com la perimetral (firewalls, Web Application Firewalls i semblants) i la local (antivirus, EDR i semblants), però és important entendre i tenir en compte les capacitats d'aquesta tecnología, sobretot amb la proliferació de ransomwares i la migració a aplicacions SaaS i entorns Cloud que hi està havent.
Amb la microsegmentació no només podem definir de manera molt granular quins moviments laterals i comunicacions estan permesos des de cada dispositiu, si no que ens permet tenir un control i una visibilitat casi total del què està passant dintre l'organització, cosa que avui en dia no es troba casi enlloc.
És per això que poder veure exemplificat les capacitats i millores de seguretat que la microsegmentació proporciona és el principi per tenir un entorn el màxim segur possible, tot seguint un model de seguretat Zero Trust.

Referències

[1] https://www2.deloitte.com/se/sv/pages/risk/articles/infrastructure-protection.html [Online; accés el 09/03/2024]

[2] https://www.ibm.com/es-es/topics/zero-trust [Online; accés el 17/03/2024]

[3] https://www.coherentmarketinsights.com/market-insight/microsegmentation-market-4235 [Online; accés el 22/06/2024]

[4] https://www.zscaler.es/resources/infographics/brief-history-zero-trust.pdf [Online; accés el 23/06/2024]

[5] https://www.netskope.com/es/security-defined/que-es-sase [Online; accés el 23/06/2024]

[6] https://www.akamai.com/es/glossary/what-is-microsegmentation [Online; accés el 09/03/2024]

[7] https://www.cybertalk.org/2023/04/06/what-is-an-island-hopping-attack-and-how-to-stop-one/ [Online; accés el 17/03/2024]

[8] https://www.crowdstrike.com/cybersecurity-101/living-off-the-land-attacks-lotl/ [Online; accés el 17/03/2024]

[9] https://www.cisco.com/site/us/en/products/security/secure-workload/index.html [Online: accés el 23/06/2024]

[10] https://www.akamai.com/es/products/akamai-guardicore-segmentation [Online; accés el 10/03/2024]

[11] https://www.illumio.com/solutions [Online; accés el 30/06/2024]

[12] https://colortokens.com/solutions/xassure-managed-microsegmentation-and-monitoring/ [Online; accés el 30/06/2024]

[13] https://www.akamai.com/infectionmonkey [Online; accés el 10/03/2024]

[14] https://github.com/guardicore/monkey [Release v2.3.0]

APÈNDIX

A.1 Figures

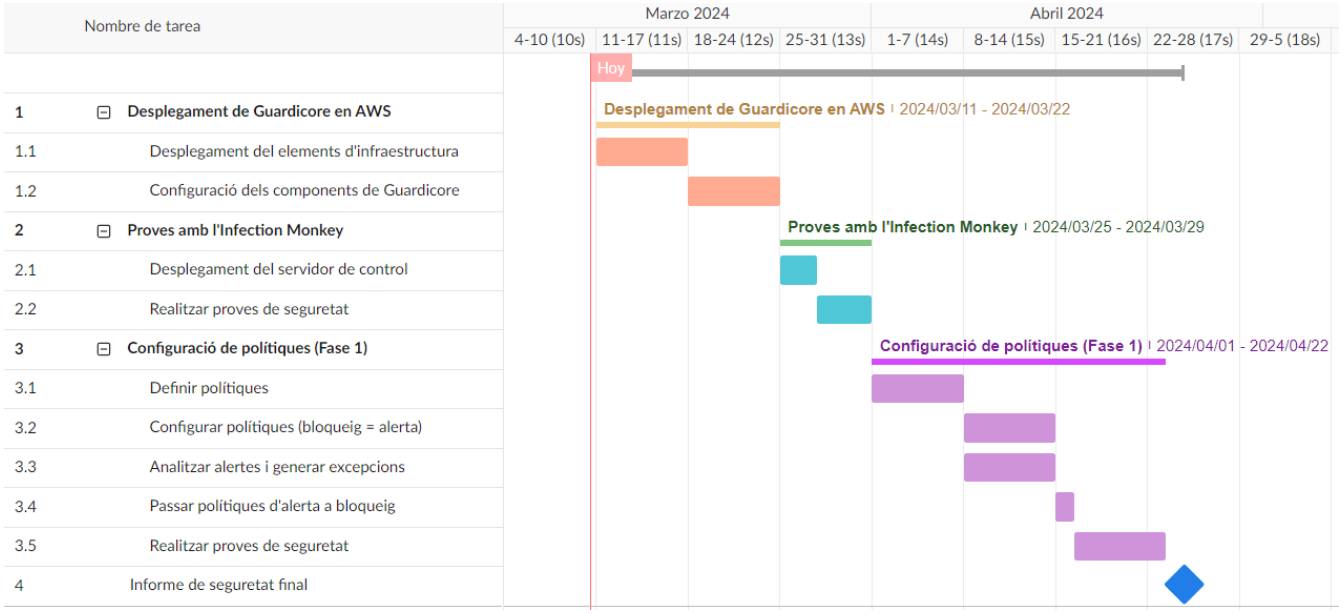

Fig. 13: *Diagrama de Gantt del projecte*

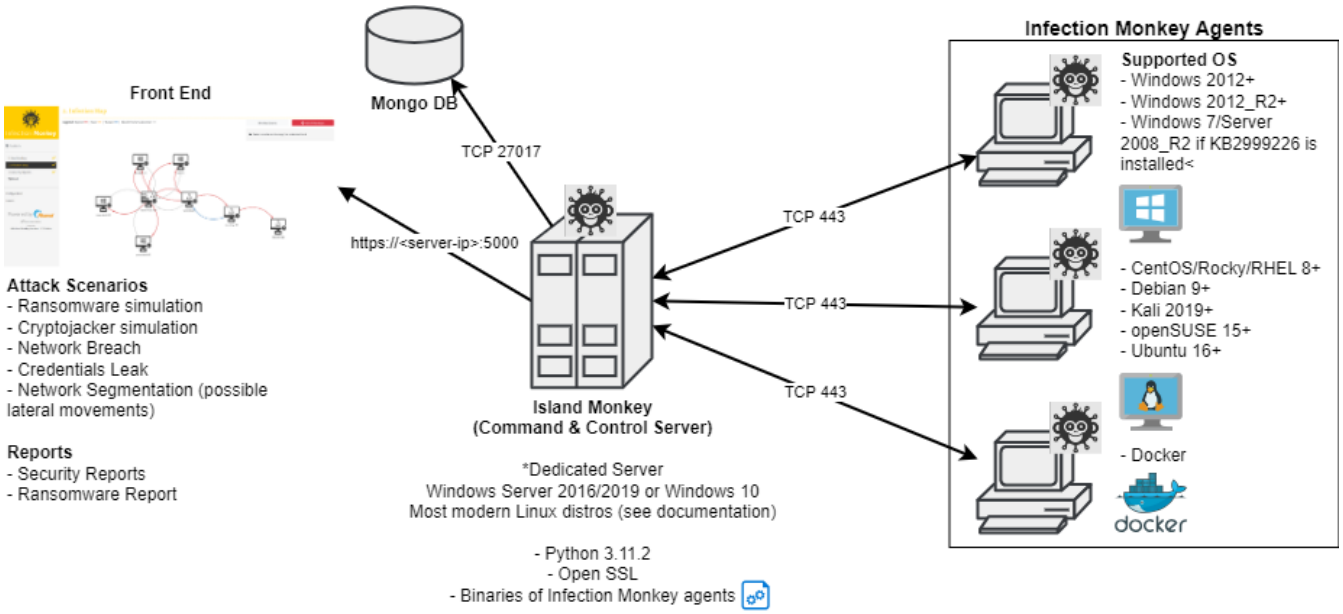

Fig. 14: *Arquitetura i comunicacions del Infection Monkey*

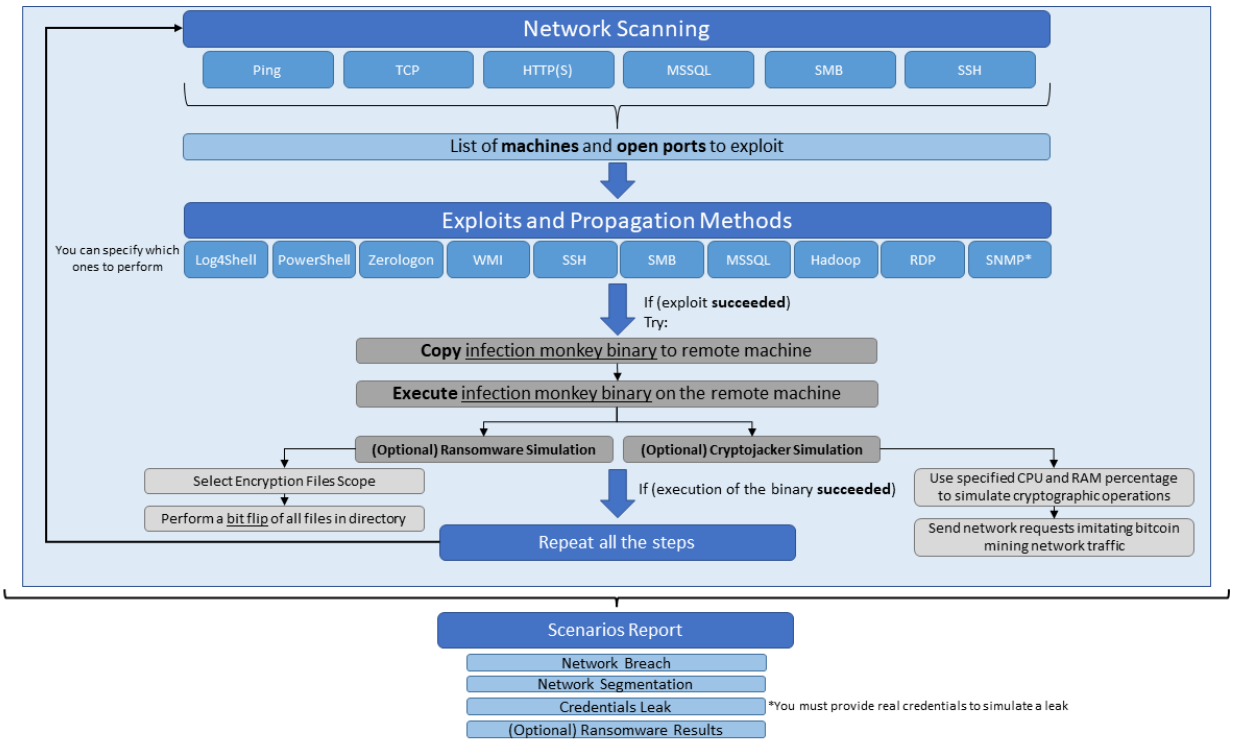

Fig. 15: *Diagrama del procés d'atac mitjançant d'exploits i simulacions del Infection Monkey*