Agent-based Modeling and Simulation for Malware Spreading in D2D Networks

Ziyad Benomar
Orange Labs
Châtillon, France
ziyad.benomar@orange.com

Chaima Ghribi
Orange Labs
Châtillon, France
chaima.ghribi@orange.com

Elie Cali
Orange Labs
Châtillon, France
elie.cali@orange.com

Alexander Hinsen
Weierstrass Institute for Applied Analysis and Stochastics
Berlin, Germany
hinsen@wias-berlin.de

Benedikt Jahnel
Weierstrass Institute for Applied Analysis and Stochastics
Berlin, Germany
jahnel@wias-berlin.de

ABSTRACT

This paper presents a new multi-agent model for simulating malware propagation in device-to-device (D2D) 5G networks. This model allows to understand and analyze mobile malware-spreading dynamics in such highly dynamical networks. Additionally, we present a theoretical study to validate and benchmark our proposed approach for some basic scenarios that are less complicated to model mathematically and also to highlight the key parameters of the model. Our simulations identify critical thresholds for "no propagation" and for "maximum malware propagation" and make predictions on the malware-spread velocity as well as device-infection rates. To the best of our knowledge, this paper is the first study applying agent-based simulations for malware propagation in D2D.

KEYWORDS

Agent-based modeling and simulation; malware spread; device-to-device

ACM Reference Format:
Ziyad Benomar, Chaima Ghribi, Elie Cali, Alexander Hinsen, and Benedikt Jahnel. 2022. Agent-based Modeling and Simulation for Malware Spreading in D2D Networks. In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), Online, May 9–13, 2022*, IFAAMAS, 9 pages.

1 INTRODUCTION

D2D communications is one of the key emerging technologies for 5G networks and beyond. It enables a direct exchange of data between mobile devices, which extends coverage for devices lacking direct access to the cellular infrastructure and therefore enhances the network capacity. However, security issues are very challenging for D2D systems as malware can easily compromise mobile devices and propagate across the decentralized network. Compromised devices represent infection threats for all of their connected neighbors as they can, in their turn, propagate malware through susceptible devices and form an epidemic outbreak. This enables attackers to infect a larger population of devices and to launch cyber- and physical malicious attacks. Therefore, it is of great importance to have a good understanding of vulnerability and security issues, particularly of the malware propagation processes, in such networks and to be able to design optimal defense strategies.

Modeling malware propagation in D2D is challenging due to the complexity of such networks induced for example by topology or device mobility. In order to cope with this, D2D can be investigated and analyzed using analytical models (e.g., stochastic geometry, stochastic processes, etc.). Some of these approaches have been proposed to model malware spreading in D2D networks [11, 20, 21]. Nevertheless, classical simulation and analytical tools are often not suitable for capturing the global dynamics of complex systems.

In this paper we propose to tackle the problem from the perspective of complex-systems science and present a new agent-based model (ABM) in order to analyze and understand malware propagation in D2D networks. For this, the agent-based simulation approach provides the possibility to simulate complex-systems dynamics and to test theories about local behaviors and their emergence. Unlike traditional techniques of simulation, based on mathematical or stochastic models, multi-agent simulation is more suitable for complex problem modeling and simulation. In fact, applying classical simulation and analytical tools, such as differential equations, to complex systems often produces undesired complications. Indeed, many challenges that arise in the traditional numerical modeling come from the fact that individual actions (activities that result in a modification of the system) and their impact on the dynamics of the system are often underrepresented. Usually, individual behaviors, i.e., decisions made at the individual or group level, cannot be incorporated into these simulations. On the other hand, in a multi-agent simulation, the model is not a set of equations as in mathematical models, but a set of entities. Here agents represent the set of all the simulated individuals, objects encode the set of all represented passive entities, and the environment is the topological space where agents and objects are located and which they can move in and act upon.

Although agent-based simulations have been successfully used to model complex systems in different areas like biology, sociology, political science and economics, it is still insufficiently explored in the field of telecommunication networks, specifically for malware spreading in D2D. In this work, we aim to shed more light on

Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), P. Faliszewski, V. Mascardi, C. Pelachaud, M.E. Taylor (eds.), May 9–13, 2022, Online. © 2022 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

whether such highly dynamical D2D networks can be treated as a complex system and whether complex-systems science can give insights on the emergent properties of malware propagation. The main contributions of this paper are as follows:

- We propose in Sections 3 and 4 a new ABM for studying malware propagation in D2D 5G+ networks and we formally prove its correctness for predicting different agents status over the time.
- We perform a theoretical study in Section 5 to estimate the critical values of the model's parameters and to identify the most important ones to consider for simulations.
- Finally, in Section 6 we perform simulations to study and understand malware-spreading dynamics. Some critical thresholds have been identified. Important aspects like malware infection rates and velocities have been also studied to understand how they will evolve as functions of the parameters.

2 RELATED WORK

ABMs are effective and robust tools in simulating complex and dynamic phenomena like epidemic spreading. These models have been used primarily in epidemiological studies of infectious diseases and have recently gained a great importance also in the epidemiological modeling as can be seen from the vast literature in the context of the COVID-19 pandemic, see for example [2, 6, 15, 16].

However, ABMs are still in their infancy with regard to telecommunication networks. Some ABMs have been proposed in the literature for IoT networks. Authors in [4, 13, 19] proposed ABMs for analyzing IoT systems. Other applications of ABMs to telecommunication networks are proposed in [17] and [18], where authors analyzed the effectiveness of ABMs to understand self-organization in peer-to peer and ad-hoc networks. These studies provide further motivation to our investigation on applying ABMs for studying malware spreading dynamics in D2D 5G networks.

Let us again mention that conventionally D2D systems are modeled using analytical methods (e.g., stochastic geometry) which have proven to be powerful tools for modeling spatial device and road systems. In this context, the authors in [20] and [21] present a framework for the modeling and understanding of malware spread in D2D with mobile devices and study some strategies of both defenders and attackers. The proposed model is based on an analytical approach and does not consider urban environments. In view of this, a standard SIR model is presented in [12], to study malware propagation in D2D considering urban environments but mobility was not taken into account. Even though the obtained results were promising, some questions remained open regarding the convergence of the malware propagation speed, the shape theorem of the infection and the critical thresholds. This mainly comes from the fact that the dynamics of the system were insufficiently captured.

3 SYSTEM MODEL

This section gives a detailed description of the D2D malware propagation model in urban environments. In this ABM description, devices are represented as reactive agents that move in the environment and have a variety of capabilities like neighborhood discovery and malware propagation. In short terms, the system has the following composition. We consider an urban environment. At initial time, devices are placed randomly on the streets (we make the simplifying assumption that devices that are situated in buildings are not to be taken into account : this can be justified by the high frequencies used in 5G). The devices move independently and randomly at a constant speed. Moreover, two devices can communicate directly with each other if they are close enough and on the same street. Let us note that this approach takes shadowing into account, but not interference. At time zero, a virus is introduced carried by a device near to the center of the city. The virus can now propagate from one device to another if they can communicate for a sufficiently long time that represents the discovery time plus the transmission time.

3.1 Street systems and devices

We consider our urban street environment E as a two-dimensional planar Poisson–Voronoi tessellation (PVT, see [5]) induced by a homogeneous Poisson point process (PPP) X_E of positive intensity λ. The PVT is a one-parameter segment process that has been shown to be a good fit for the street systems of European cities (see [8–10]). It has been widely used to model different urban environments as random tessellations , since it allows to go beyond specific urban topologies. We will denote by S the set of edges of E (representing the streets). The devices are placed on S as a linear PPP of intensity θ, thus forming a Cox point process (CPP) on the plane with random intensity measure $\Lambda(B) = \theta|S \cap B|$ for every measurable $B \in \mathbb{R}^2$. Here $|S \cap B|$ stands for the total length of S in the area B.

3.2 ABM for malware propagation in D2D

We note first that the environment is modeled as an undirected graph, relying on some stochastic-geometry concepts, as described in Section 3.1. Then, we define our malware-propagation system in D2D as a finite number of agents, states, actions and rules,

$$\text{MAS} := \langle \mathcal{A}, \text{St}, \text{Act}, \text{R}, \text{T} \rangle.$$

More precisely, we consider a set of n *agents* $\mathcal{A} = \{ a_i : i \in [1, n] \}$ corresponding to devices and a state space St = {susceptible, infected}. Further, Act = {"move", "discover", "connect", "infect"} denotes the set of *actions* that each agent can perform according to its state. R represents the set of the *behavioral rule base.* Time T is assumed to be divided in time units called *slots*, where each slot k is represented by a positive integer.

Initially, agents are distributed on the edges of E (i.e., streets of the city) as described in Section 3.1. One agent of type *infected* is introduced around the center of the map. Then, formally, each agent a_i is defined at each time slot by a tuple

$$\mathcal{M}_{i,k} := \langle X_{i,k}, V_{i,k}, N_{i,k}, \text{Act}_{i,k}, \xi_{i,k}, T_{i,k}^{(I)} \rangle.$$

Here, $X_{i,k}$ specifies the agent's *location* in terms of coordinates at time kdt, $V_{i,k} = v$ represents the agent's moving *speed* and $N_{i,k}$ the *knowledge base*, representing what each agent a_i knows about its neighboring agents and the environment at time slot k. Further, $\xi_{i,k} \in$ St represents the state of agent a_i and $\text{Act}_{i,k}$ is the set of actions that can be performed by a_i. Finally, $T_{i,k}^{(I)}$ represents the *first time* when a_i becomes infected. It will be updated during the simulation depending on the agent's interactions. $T_{i,0}^{(I)}$ is set to $+\infty$ for initially susceptible agents and 0 for the infected one. The state

of $a_i \in \mathcal{A}$ at kdt for $k \geq 1$ is given by

$$\xi_{i,k} := \begin{cases} \text{susceptible} & \text{if } kdt < T^{(I)}_{i,k-1}, \\ \text{infected} & \text{if } kdt \geq T^{(I)}_{i,k-1}. \end{cases}$$

In particular, the state of a_i at a step k of the simulation is computed using the variables $(T^{(I)}_{i,k-1})_{a_i \in \mathcal{A}}$ from the previous step. This formula also implies that the states of the agents will not change between the steps 0 and 1. It will be indeed the case since we will consider a time step dt smaller than ρ. (See Section 4.2).

3.3 Agent behavior and states

Let us describe the three different behaviors of agents: mobility, communication and infection.

3.3.1 Mobility behavior. Devices move at the same constant speed v, starting from a base position, and repeating indefinitely the following *street-adapted random-waypoint model*:

- Each device independently picks a destination on the street. For this we sample a random point P in the plane using a Gaussian distribution centered on the device X, and with a standard deviation equal to $\sigma_X = (15\text{min}) \times v$. The *destination* we take for X is then the closest point of P in E. This choice of σ_X shows that devices will go to destinations that they can reach in an average time of 15min if they take a straight path.
- Devices move to their destinations following the shortest path along the streets.
- Once arrived, devices go back to their starting position following the shortest path along the streets (anchored movement).

3.3.2 Communication behavior. In order to exchange messages, two communicating devices/agents must obey the following rules:

- (RAD): The Euclidean distance between the two devices is less than a given constant threshold r.
- (LOS): The two devices are on the same street.

The first rule supposes that the emission power of the devices is a constant and that we do not take into account interference. The second rule means that the signal cannot go through the buildings and that reflections and diffractions are not taken into account. In symbols, for $X_i(t)$ the position of device a_i at time t and $\mathcal{N}(a_i, a_j) := \{t \geq 0 \colon \|X_i(t) - X_j(t)\| < r \text{ and } \exists s \in S \text{ such that } (X_i(t), X_j(t)) \in s\}$, we have that a_i and a_j are connected at time t if and only if $t \in \mathcal{N}(a_i, a_j)$.

3.3.3 Infection behavior. We will follow a standard SI compartmental model, very similar to SIR which is a classical approach in epidemiology often used within the framework of differential equations. However, unlike the latter, in a D2D context, users are constrained to be positioned on streets and are mobile, two aspects that are usually not represented in epidemiological studies. The SI model is formulated by first partitioning devices into two distinct categories called susceptible (S) and infected (I). At time zero, only one device will be in the infected state, while a CPP X_S with intensity θ will define the susceptible devices, independent of the former one given the PVT tessellation. When an infected device is connected to a susceptible device for a time longer than a given threshold ρ, the susceptible device will become infected. More precisely, if the device a_i is infected at time t and if $[t, t+\rho] \in \mathcal{N}(a_i, a_j)$, then a_j is infected at $(t + \rho)$.

3.3.4 Agent states. Agent states specify what state an agent is in. Agent-state transitions are driven by the rule base R that implements the reactive behavior of agents. It allows to select actions to take for agent a_i depending on its current local state $\xi_{i,k}$ and its knowledge base $N_{i,k}$. More specifically, we write R = $\{\Theta\}$ where $\Theta(\xi_{i,k}, N_{i,k})$ are the *active rules* that map the set of states and observations to actions for reactive tasks

$$\Theta : (\xi_{i,k}, N_{i,k}) \longrightarrow \text{Act}_{i,k}.$$

Let $T^{(C)}_{i,j}$ be the connection duration between agents a_i and a_j and ρ be the needed time for the virus transmission from one agent to another. Then the principal rule-based function is described as follows.

- Malware infection rule: If agent a_i is infected, agent a_j is susceptible ($\xi_{i,k}$ = infected, $\xi_{j,k}$ = susceptible) and a_i was connected to a_j for a time longer than the infection threshold ($T^{(C)}_{i,j} \geq \rho$), then the state of agent a_j will be transited from susceptible to infected (the action infect will be activated),

$$\Theta_I : (\xi_{i,k}, N_{i,k}) \longrightarrow \textbf{Infect}.$$

A more detailed description of the algorithm associated to malware infection will be given in Section 4.

4 AGENT-BASED SIMULATION

In this section we present more details on the implementation of our multi-agent simulation tool. Let us denote by

$$\mathcal{P} := \{dt, \rho, r, \lambda, \theta, v\}$$

the set of key model parameters where dt represents the elapsed time in each step, ρ and r represent respectively connection time needed for virus transmission and communication radius of agents. λ is the intensity of Voronoi seeds (seed/km^2), θ is the intensity of susceptible agents (agent/km) and v denotes agents speed (km/h). Other parameters such as the dimensions (H_1, H_2) of the map can be added to this list, but we will not focus on these in our study. For the most part of the manuscript, we give the same speed to all the agents in order to keep a restraint number of parameters. However, we can easily have a more general model where the speeds of the agents are distributed following some probability law. Each agent could have for example a speed taken uniformly at random in some interval $[v_1, v_2]$.

Our simulation is done over steps, each step corresponds to a time instant kdt. In the following we will denote by $\mathcal{M}_k$ the model at step k. It represents the map, the agents and all their attributes (coordinates, states, etc.) at step k. In the simulation, we first generate a random map, then the agents, and after that we run the function Step($\mathcal{M}_k$), that updates the variables of the model, taking it from a step k to the next step $k + 1$, for a number $k_{\max}$ of iterations. Algorithm 1 describes the entry function of the simulation. The function GenerateMap(λ) returns a random PVT with parameter λ, whereas the function Generate(θ, v) returns the set of agents $\mathcal{A} := \mathcal{A}_S \cup \{a_{i_0}\}$, where $\mathcal{A}_S$ is the set of initially susceptible agents distributed on $\mathbb{M}$ using homogeneous PPP with parameters θ. a_{i_0} is the initially infected agent, placed near the center of the map.

Algorithm 1: Main($\mathcal{P}, k_{\max}$): The main function describing the simulation

Input : The set of parameters $\mathcal{P}$ and the maximum number of steps $k_{\max}$
Output: The state of a randomly generated model at time $k_{\max}dt$

$\mathbb{M} \leftarrow$ GenerateMap(λ);
$\mathcal{A} \leftarrow$ GenerateAgents(θ);
$\mathcal{M}_0 \leftarrow (\mathcal{P}, \mathbb{M}, \mathcal{A}, (X_{i,0})_i, (T^{(I)}_{i,0})_i)$;
for $k \in \{1, \dots, k_{\max}\}$ **do**
  $\mathcal{M}_k \leftarrow$ Step($\mathcal{M}_{k-1}$);
return $\mathcal{M}_{k_{\max}}$

4.1 Discrete-time approximations

Recall that our simulations are done over steps, where each step k corresponds to a time instant kdt. A difficulty lies in the correct updating of the states of the agents. From step k to $k+1$, each agent moves independently as described in Section 3.3.1, which means that we can access the positions of the agents at times $(kdt)_{k\in\mathbb{N}}$ knowing their velocities and the edges they have been through, but it is complicated to know all the interactions they had given only this information. To overcome this, we first impose the constraint $dt < \rho$. This guarantees that, by only observing the positions of the agents at discrete times with a step dt, we will not miss any two devices that connect for a duration longer than ρ, see Section 4.2 for more details. Let $k \in \mathbb{N}$, and let us assume that a_i, a_j are connected to each other at kdt. We will treat the general case where they can have different speeds v_i and v_j, and we will compute the duration of the connection using their movement equations. Let us denote by $t^{(\text{in})}_{i,s}$ (respectively $t^{(\text{out})}_{i,s}$) the time when a_i gets in (respectively out) of the street s. These can easily be computed knowing X_i and the length $L(s)$ of the street s. Since s has two different directions, we need to consider their velocities $\mathbf{v}_i, \mathbf{v}_j$. Let P_1, P_2 be the positions of the two extremities of the street s, let $\mathbf{e} := (P_2 - P_1)/\|P_2 - P_1\|$ (we can take $-\mathbf{e}$ instead), and v_i, v_j be such that $\mathbf{v}_i = v_i\mathbf{e}, \mathbf{v}_j = v_j\mathbf{e}$. We recall that the absolute speed v_i of a_i obeys $v_i = \|\mathbf{v}_i\| = \pm v_i$, the same holds for a_j. Finally, let us also define the coordinates of a_i, a_j on the street s by $d_{i,k} := (X_{i,k} - P_1) \cdot \mathbf{e}$ and $d_{j,k} := (X_{j,k} - P_1) \cdot \mathbf{e}$. Then we have the following result that we present without proof.

Lemma 1. *If a_i, a_j are connected at time kdt and if $v_i \neq v_j$, then they are connected during all the time interval $[t^{(C,i)}_{i,j}, t^{(C,f)}_{i,j}]$, where*

$$t^{(C,i)}_{i,j} := \max\{kdt - \frac{d_{i,k} - d_{j,k}}{v_i - v_j} - \frac{r}{|v_i - v_j|}, t^{(in)}_{i,s}, t^{(in)}_{j,s}\},$$

$$t^{(C,f)}_{i,j} := \min\{kdt - \frac{d_{i,k} - d_{j,k}}{v_i - v_j} + \frac{r}{|v_i - v_j|}, t^{(out)}_{i,s}, t^{(out)}_{j,s}\}.$$

Moreover, if $v_i = v_j$, then

$$t^{(C,i)}_{i,j} = \max\{t^{(in)}_{i,s}, t^{(in)}_{j,s}\} \text{ and } t^{(C,f)}_{i,j} = \min\{t^{(out)}_{i,s}, t^{(out)}_{j,s}\}.$$

The connection duration of a_i, a_j is then $T^{(C)}_{i,j} := t^{(C,f)}_{i,j} - t^{(C,i)}_{i,j}$.

In words, two agents on the same street can have different speeds and move either in the same or in opposite directions. Recall that the connection-time interval is the set of all time instants such that the distance of the two agents is less than r.

We saw in Section 3.2, that agents states will be determined by the variable $T^{(I)}_{i,k-1}$ at each step $k \geq 1$. We call $\mathcal{S}_k, \mathcal{I}_k$ the sets of susceptible and infected agents. Let ConnectionInterval(a_i, a_j, k) be a function computing $t^{(C,i)}_{i,j}, t^{(C,f)}_{i,j}$ as in Lemma 1 knowing that a_i, a_j are connected at kdt, and let GetNeighbors(a_i) be a function returning the set of neighbors of a_i defined as: $N_k(a_i) := \{a_j \in \mathcal{A} : \|X_{i,k} - X_{j,k}\| \leq r \text{ and } a_i, a_j \text{ on the same street}\}$. Finding the neighbors of all the agents would normally require a $O(n^2)$ time complexity, but since only agents on the same street can connect to each other, we can considerably reduce this complexity by searching neighbors of each agent only among those that are on the same street. From here, we can write Algorithm 2 that updates the values $T^{(I)}_{j,k}$ for the neighbors of an infected agent a_i.

Algorithm 2: InfectNeighbors(a_i)

Input : An infected agent a_i
Output: Updates $T^{(I)}_{j,k}$ for all susceptible neighbors of a_i

1 $N^{(S)}_k(a_i) \leftarrow$ GetNeighbors(a_i) $\cap \mathcal{S}_k$;
2 **for** $a_j \in N^{(S)}_k(a_i)$ **do**
3 $t^{(C,i)}_{i,j}, t^{(C,f)}_{i,j} \leftarrow$ ConnectionInterval(a_i, a_j, k);
4 $t^{(C,i)}_{i,j} \leftarrow \max\{t^{(C,i)}_{i,j}, T^{(I)}_{i,k}\}$;
5 **if** $T^{(C)}_{i,j} := t^{(C,f)}_{i,j} - t^{(C,i)}_{i,j} \geq \rho$ **then**
6 $T^{(I)}_{j,k} \leftarrow \min\{T^{(I)}_{i,k}, t^{(C,i)}_{i,j} + \rho\}$;

Line 4 makes sure that we only compute the time when the agents are connected and a_i is infected. Note that in Line 6, we cannot set the value of $T^{(I)}_{j,k}$ simply to $t^{(C,i)}_{i,j} + \rho$ as agent a_j might be connected to several infected agents, and it will become infected as soon as it stays connected to one of them for longer than ρ. Finally, we can write the core function of our simulation, that is Algorithm 3.

Algorithm 3: The Step Function

Input : The model $\mathcal{M}_{k-1}$ at step $k-1$
Output: The model $\mathcal{M}_k$ at step k

$\mathcal{S}_k, \mathcal{I}_k \leftarrow$ The sets of susceptible and infected agents;
for $a_i \in \mathcal{A}$ **do**
  $X_{i,k} \leftarrow$ Move($a_i, V_i, X_{i,k-1}, dt$); //Update the positions
  $T^{(I)}_{i,k} \leftarrow T^{(I)}_{i,k-1}$; //Initialisation
for $a_i \in \mathcal{I}_k$ **do**
  InfectNeighbors(a_i); //Update the variables $T^{(I)}_{j,k}$
$\mathcal{M}_k \leftarrow (\mathcal{P}, \mathbb{M}, \mathcal{A}, (X_{i,k})_i, (T^{(I)}_{i,k})_i)$;
return $\mathcal{M}_k$;

4.2 Equivalence of discrete and continuous time

We denote by $\xi_i(t)$ the state of agent a_i at *continuous time* t for any $a_i \in \mathcal{A}$. On the other hand, for each $k \in \mathbb{N}$ we denote as before by $\xi_{i,k}$ the state of a_i at *discrete time* kdt as predicted by our ABM. The following theorem states that for sufficiently small time slots, at the discrete time points, our model is equivalent to its continuous-time version and is then theoretically proven to be correct.

Theorem 2. *If $dt < \rho$, then we have*

$$\forall a_i \in \mathcal{A}, \forall k \in \mathbb{N}, \qquad \xi_{i,k} = \xi_i(kdt).$$

In words, Theorem 2 guarantees that, by discretizing, we do not miss infection events and the introduced time differences do not induce errors in the discretized model. Let us first define the first continuous time when $a_j \in \mathcal{A}$ is infected, i.e., $\tilde{T}_j^{(I)} := \inf\{t \geq 0 \colon \xi_j(t) = \text{infected}\}$. Regarding our malware propagation rules, we can write

$$\tilde{T}_j^{(I)} = \inf_{a_i \neq a_j} \inf_{t \geq \tilde{T}_i^{(I)}} \{t + \rho \colon [t, t+\rho] \subset \mathcal{N}(a_i, a_j)\}, \tag{1}$$

where $\mathcal{N}(a_i, a_j)$ is as defined in Section 3.3.2. Let us also denote $\mathcal{S}_k := \{a_i \colon kdt < T_{i,k-1}^{(I)}\}$, $\tilde{\mathcal{S}}_k := \{a_i \colon kdt < \tilde{T}_i^{(I)}\}$, $\mathcal{I}_k := \{a_i \colon kdt \geq T_{i,k-1}^{(I)}\}$ and $\tilde{\mathcal{I}}_k := \{a_i \colon kdt \geq \tilde{T}_i^{(I)}\}$. Finally, for convenience, let $T_{i,-1}^{(I)} := T_{i,0}^{(I)}$ for all $a_i \in \mathcal{A}$. We have the following lemma.

Lemma 3. *If $dt < \rho$, then for any $k \in \mathbb{N}$, assertion $\mathcal{B}_k$ is true*

$$(\mathcal{B}_k) \ : \qquad \forall a_j \in \mathcal{A}, \ \begin{cases} \tilde{T}_j^{(I)} \leq T_{j,k-1}^{(I)}, \\ \tilde{T}_j^{(I)} \leq kdt \implies \tilde{T}_j^{(I)} = T_{j,k-1}^{(I)}. \end{cases}$$

Note that, if $\mathcal{B}_k$ is verified for some $k \in \mathbb{N}$, then $\tilde{\mathcal{S}}_k \subset \mathcal{S}_k$ and $\tilde{\mathcal{I}}_k \subset \mathcal{I}_k$. But since $\tilde{\mathcal{S}} \cup \tilde{\mathcal{I}}_k = \mathcal{S}_k \cup \mathcal{I}_k$, this means that $\tilde{\mathcal{S}}_k = \mathcal{S}_k$ and $\tilde{\mathcal{I}}_k = \mathcal{I}_k$ and thus Theorem 2 is proved.

Proof. For $k = 0$ the assertion is true by definition of $(T_{i,-1}^{(I)})_{a_i \in \mathcal{A}}$. Let $k \geq 1$, assume that $\mathcal{B}_k$ is true and let $a_j \in \mathcal{A}$. If $T_{j,k}^{(I)} = T_{j,k-1}^{(I)}$ then directly $\tilde{T}_j^{(I)} \leq T_{j,k}^{(I)}$. Otherwise $T_{j,k}^{(I)}$ was updated during step k, i.e., there exists an agent $a_i \in \mathcal{I}_k$ for which InfectNeighbor(a_i) was called and such that $a_j \in N_{i,k}^{(S)}$ and $t_2 - t_1 \geq \rho$ with $t_1 = \max\{t_{i,j}^{(C,i)}, T_{i,k-1}^{(I)}\}$, $t_2 = t_{i,j}^{(C,f)}$. This implies that $[t_1, t_1 + \rho] \subset [t_{i,j}^{(C,i)}, t_{i,j}^{(C,f)}] \subset \mathcal{N}(a_i, a_j)$, and since $a_i \in \mathcal{I}_k$, we have by the induction hypothesis that $\tilde{T}_i^{(I)} = T_{i,k-1}^{(I)}$ and thus $t_1 \geq \tilde{T}_i^{(I)}$. Thus, using Equation (1), we have that $\tilde{T}_j^{(I)} \leq T_{j,k}^{(I)}$. For the second part of the assertion, let us assume that $\tilde{T}_j^{(I)} \leq (k+1)dt$. If $\tilde{T}_j^{(I)} \leq kdt$ then $T_{j,k}^{(I)} \leq T_{j,k-1}^{(I)} = \tilde{T}_j^{(I)}$ (induction hypothesis), and we proved that $\tilde{T}_j^{(I)} \leq T_{j,k}^{(I)}$ and therefore $\tilde{T}_j^{(I)} = T_{j,k}^{(I)}$. Otherwise $kdt < \tilde{T}_j^{(I)} \leq (k+1)dt$, this implies that $a_j \in \tilde{\mathcal{S}}_k$ and there exists $a_i \in \mathcal{A}$ such that $[\tilde{t}, \tilde{t} + \rho] \subset \mathcal{N}(a_i, a_j)$ and $\tilde{t} \geq \tilde{T}_i^{(I)}$ with $\tilde{t} := \tilde{T}_j^{(I)} - \rho$. Given that $dt < \rho$ we have

$$\tilde{T}_i^{(I)} \leq \tilde{t} = \tilde{T}_j^{(I)} - \rho_I \leq (k+1)dt - \rho_I < kdt < \tilde{T}_j^{(I)} = \tilde{t} + \rho_I,$$

and this implies that $a_i \in \tilde{\mathcal{I}}_k$ and $kdt \in [\tilde{t}, \tilde{t} + \rho] \subset \mathcal{N}(a_i, a_j)$. Thus InfectNeighbors is called on a_i at step k and a_j is among the visited agents during this call (neighbors of a_i). $\tilde{T}_{j,k}^{(I)}$ will then be updated and its final value will be at most $\tilde{t} + \rho = \tilde{T}_i^{(I)}$. With the inequality $\tilde{T}_j^{(I)} \leq T_{j,k}^{(I)}$ that we already proved, we deduce that $\tilde{T}_j^{(I)} = T_{j,k}^{(I)}$. □

Finally, for any $k \in \mathbb{N}$, we have by Lemma 3 that $\tilde{\mathcal{S}}_k = \mathcal{S}_k$ and $\tilde{\mathcal{I}}_k = \mathcal{I}_k$. This means that the states of the agents predicted by the simulator correspond to their real states.

5 MEAN-FIELD VERSION

The model that we presented so far is very rich with many parameters. It is therefore difficult to run simulations varying all these parameters and see how each of them influences the propagation of the virus. So, in order to better choose the values we will assign to them, in this section, we present a theoretical study on a simplified model to identify critical relationships between parameters and values that will lead to drastic changes in the system's evolution. Let us highlight that we consider a different model that does not arise as a limiting object. It is mainly introduced in order to sharpen the intuition for threshold values of important parameters.

As in the first model, we start with a single infected agent a_{i_0}, and we will take interest in the time of the first virus transmission, which we will denote by τ in the following. Let us stress that the simplified model that we present here is used only as a mathematical model. All the simulations results in Section 6 are based on the original model and not this simplified one.

We consider the following mean-field approximation of our spatial model. Instead of considering a_{i_0} to be moving on a PVT, we will consider that it moves on a succession of streets $s_0, s_1, \ldots$, each having a length $L_\lambda^{(i)}$ that is a random variable with density $f_{\lambda,L}$, where $f_{\lambda,L}$ is the density function of the edges lengths in a PVT having a seeds intensity equal to λ (see Section 3.1). We will assume that, when a_{i_0} enters a street, other agents are distributed on it as an homogeneous PPP with parameter θ, and that they can move in any of the two possible directions. What we mainly lose in this simplified model is the dependence between the lengths of the successive streets visited by a_{i_0}.

For each street s_i visited by a_{i_0}, let C_i be the number of agents that a_{i_0} infects while being on s_i. Let $p := \Pr[C_i \geq 1]$ denote the probability that a_{i_0} infects at least some agent on s_i (p is independent of i). Finally, let τ be the first time instant when some agent a_j different from a_{i_0} becomes infected

$$\tau := \inf\{t \geq 0 \colon \exists j \neq i_0 \text{ such that } \xi_j(t) = \text{infected}\}.$$

Then, we have the following main results.

Theorem 4. *If τ is the first time when a_{i_0} infects another agent, then*

$$\frac{2}{3\sqrt{\lambda}v}(1/p - 1) \leq \mathbb{E}[\tau] \leq \frac{2}{3\sqrt{\lambda}v} \cdot 1/p.$$

Theorem 5. *There exists a positive constant $\tilde{C}$ such that if p is sufficiently small, then for $t_0 = 1/(3\sqrt{p\lambda}v)$ we have*

$$\Pr[\tau \geq t_0] \geq 1 - \tilde{C}p^{1/4}.$$

These theorems indicate that, if the probability of infecting another agent on a single street s_i is low, then the waiting time before the virus transmission is very large, and therefore the virus propagation is weak. In terms of the asymptotic behavior of the system, we can state that, when $p = o(1)$, then $\mathbb{E}[\tau] = \Omega(1/(\sqrt{\lambda}vp)$ and for $t_0 = 1/(3\sqrt{p\lambda}v)$ we have $\Pr[\tau \geq t_0] = 1 - O(p^{1/4})$.

The proofs rely on results for typical edge length in PVT and Berry–Esseen inequalities. Let us start by presenting a first lemma on the edges-lengths distribution when $\lambda = 1$, as described in [3].

Lemma 6. *In a random planar PVT, if we choose a random edge, then its length L is a random variable having a distribution f_L satisfying*

(1) $f_L(0) = 2/\pi$, *when l is large enough:* $f_L(l) \sim \frac{\pi^2}{3\sqrt{2}} l^2 e^{-\frac{\pi}{2}l^2}$,
(2) *if L is a random variable with density function f_L, then L has an n-th moment for any positive integer n and*
(3) $\mathbb{E}[L] = 2/3$, $\sigma_L^2 := \mathrm{Var}[L] \approx 0.1856$.

From this we deduce the result for any positive λ.

Lemma 7. *If P_λ is a PVT generated with an intensity of seeds $\lambda > 0$, then the edges length in P_λ will have a density function $f_{\lambda,L}$ given by*

$$f_{\lambda,L}(l) := \sqrt{\lambda} f_L(\sqrt{\lambda} l), \qquad \forall l \geq 0.$$

Then we have the following statement.

Corollary 8. *If L_λ is a random variable with density $f_{\lambda,L}$, then for any positive integer n, L_λ has a n-th moment given by*

$$\mathbb{E}[L_\lambda^n] = \lambda^{-n/2} E[L_1^n].$$

Finally, since f_L is a rapidly decreasing function when ℓ is large, we have the following probability estimate.

Lemma 9. *There exists $l_0 > 0$ such that for any $x \geq l_0/\sqrt{\lambda}$*

$$\Pr[L_\lambda \geq x] \leq \exp(-\lambda x^2).$$

Next, the following theorem is a corollary of the Berry–Esseen's inequality [1, 7] applied to random variables $(X_i - \mu)$ and using the trivial relation $\Pr[Y_n > x] = 1 - \Pr[Y_n \leq x]$ for $Y_n := \frac{1}{n}\sum_{i=0}^{n-1}(X_i - \mu)$.

Theorem 10. *There exists a constant C such that if $X_0, X_1, \ldots$ are i.i.d. random variables with $\mathbb{E}[|X_0|] = \mu < +\infty$, $\mathrm{Var}[X_0] = \sigma^2 > 0$ and $\mathbb{E}[|X_0 - \mu|^3] = m < +\infty$, then for any $n \in \mathbb{N}$ and $x \in \mathbb{R}$*

$$\Pr\left[\sum_{i=0}^{n-1} X_i > x\right] \geq 1 - \Phi\left(\left(\frac{x}{n} - \mu\right)\frac{\sqrt{n}}{\sigma}\right) - \frac{Cm}{\sigma^3\sqrt{n}}.$$

Here Φ is the cumulative distribution function of the standard normal distribution.

We are now in the position to prove our main theorems.

Proof of Theorem 4. We only need to observe that $\tau_0 + \ldots + \tau_{m-1} \leq \tau \leq \tau_0 + \ldots + \tau_m$, where m is the index of the first street such that $C_m \geq 1$ and $\tau_i := L_\lambda^{(i)}/v$ is the time spent by a_{i_0} on the street i. Using the law of total expectation, we deduce that $\mathbb{E}[m]\mathbb{E}[L_\lambda]/v \leq \mathbb{E}[\tau] \leq (\mathbb{E}[m]+1)\mathbb{E}[L_\lambda]/v$, and the result is obtained by computing the two expectations $\mathbb{E}[m]$ and $\mathbb{E}[L_\lambda]$. □

Proof of Theorem 5. To prove this theorem, we first need to observe that for any $\tilde{m} \in \mathbb{N}$ we have $\Pr[\tau \geq t_0] \geq \Pr[\tau_0 + \ldots + \tau_{m-1} \geq t_0] \geq \Pr[\tau_0 + \ldots + \tau_{\tilde{m}-1} \geq t_0]\Pr[m \geq \tilde{m}]$ (the second inequality is true because all the τ_i are non-negative). In particular for $\tilde{m} = \lceil 1/\sqrt{p} \rceil \leq 1/\sqrt{p} + 1$, Bernoulli's inequality gives that

$$\Pr[m \geq \tilde{m}] = (1-p)^{\tilde{m}} \geq 1 - \tilde{m}p \geq 1 - \sqrt{p} - p \geq 1 - 2\sqrt{p}.$$

On the other hand, if $p \leq (6\sigma_L)^{-4}$, using Theorem 10 and the inequality $\Phi(-x) \leq \exp(-x)/\sqrt{2\pi}$, which is true for any $x \geq 2$, we find a constant C_1 verifying

$$\Pr[\tau_0 + \ldots + \tau_{\tilde{m}-1} \geq t_0] \geq 1 - C_1 p^{1/4}.$$

Finally, when p is small enough, we deduce using again Bernoulli's inequality that

$$\Pr[\tau \geq t_0] \geq (1 - 2\sqrt{p})(1 - C_1 p^{1/4}) \geq 1 - (2 + C_1)p^{1/4}.$$

This finishes the proof. □

We will now apply the previous theorems to show that the virus propagation is slow in any of the following cases.

- The transmission time of the virus is very large compared to the expected time spent by agents on streets : $\sqrt{\lambda}\rho v \gg 1$
- The number of agents reachable within the communication radius is very small: $\theta r \ll 1$
- The number of agents on each street is very small: $\theta/\sqrt{\lambda} \ll 1$.

Corollary 11. *If $\sqrt{\lambda}\rho v \geq l_0$, then $\mathbb{E}[\tau] \geq \dfrac{2}{3\sqrt{\lambda}v}(e^{\lambda\rho^2 v^2} - 1)$.*

Proof. We have the implications $C_0 \geq 1 \Rightarrow \tau_0 \geq \rho \Rightarrow L_\lambda^{(0)} \geq \rho v$. Hence, using Lemma 9, $p = \Pr[C_0 \geq 1] \leq \Pr[L_\lambda \geq \rho v] \leq \exp(-\lambda\rho^2 v^2)$. The result follows directly from Theorem 4. □

Corollary 12. *If $r < \rho v$, then $\mathbb{E}[\tau] \geq \dfrac{2}{3\sqrt{\lambda}v}(1/(\theta r) - 1)$.*

Proof. If $r < \rho v$, then a_{i_0} can only infect the agents moving in the same direction as him: its connection time with the agents moving in the opposite direction is upper bounded by $r/v < \rho$. Let N_c be the number of agents that a_{i_0} connects to while being on s_0. Since each agent in s_0 can be moving in any of the directions with a probability $1/2$, N_c is dominated by a random Poisson variable with parameter $2\theta r/2 = \theta r$, which means that

$$\begin{aligned}\Pr[N_c \geq 1] &= \int_0^{+\infty} \Pr[N_c \geq 1 \mid L_\lambda^{(0)} = \ell] f_{\lambda,L}(\ell) d\ell \\ &\leq \int_0^{+\infty} (1 - e^{-\theta r}) f_{\lambda,L}(\ell) d\ell = 1 - e^{-\theta r} \leq \theta r.\end{aligned}$$

Since $C_0 \leq N_c$, Theorem 4 gives the desired result. □

Corollary 13. *We have $\mathbb{E}[\tau] \geq (\sqrt{\lambda}/\theta - 4/3)/(2\sqrt{\lambda}v)$.*

Proof. We can easily prove that the number N of agents that were on s_0 at some time instant when a_{i_0} was on it too follows a Poisson distribution with parameter $2\theta L_\lambda^{(0)}$. In fact, when a_{i_0} enters s_0, there are N_1 agents on the street, and by the time it reaches the end of the street, since all the agents have the same speed, all of these will have left it and N_2 new agents will have come. N_1 and

N_2 both follow a Poisson distribution with parameter $\theta L_\lambda^{(0)}$ and $N = N_1 + N_2$. Finally, given that $C_0 \leq N$, we have

$$\Pr[C_0 \geq 1] \leq \Pr[N \geq 1] = \int_0^{+\infty} \Pr[N \geq 1 \mid L_\lambda^{(0)} = l] f_{\lambda,L}(l) dl$$
$$= \int_0^{+\infty} (1 - e^{-2\theta l}) f_{\lambda,L}(l) dl \leq \frac{2\theta}{\sqrt{\lambda}} \mathbb{E}[L_1] = \frac{4\theta}{3\sqrt{\lambda}}.$$

Applying Theorem 4 concludes the proof. □

Using Theorem 5 in these three cases, we can also find lower bounds for τ that hold with high probability.

6 SIMULATION RESULTS

This section discusses simulations that were performed to analyze malware propagation in D2D, to benchmark the mathematical study in Section 5 and to show how the various parameters accelerate or slow down the propagation. Our ABM was built based on Mesa [14], which is a very suitable python framework for ABMs that we have extended to generate and visualize street system environments.

6.1 Evaluation indicators

We present some indicators that allow us to analyze malware propagation. They should be independent of the dimensions of the map, since we theoretically want to study propagation on an infinite plan. We denote τ_u the time when the infection reaches the distance u from the initial infection point

$$\tau_u := \inf\{t \geq 0 \colon \exists a_j \in \mathcal{I}(t) : \|X_j(t) - X_{I_0}(0)\| \geq u\}.$$

where $\mathcal{I}(t)$ is the set of infected agents at time t, a_{I_0} the only initially infected agent and X_{I_0} its position at time 0.

Definition 14 (**Propagation speed**). *The propagation speed is the velocity of malware spread in space. It is defined by*

$$\mathcal{V} := \limsup_{u \to +\infty} u \mathbb{E}[1/\tau_u],$$

Definition 15 (**Infection rate**). *The infection rate is the rate of infected agents in the region reached by the virus*

$$\mathcal{R} := \limsup_{u \to +\infty} \frac{|\mathcal{I}(\tau_u)|}{|\{X_j(\tau_u) \colon a_j \in \mathcal{A}\} \cap B(X_{I_0}(0), u)|},$$

where $B(X_{I_0}(0), u)$ is the open ball of center $X_{I_0}(0)$ and radius u.

Note that $|\{X_j(\tau_u)\} \cap B(X_{I_0}(0), u)|$ is simply the number of agents inside $B(X_{I_0}(0), u)$ at time τ_u.

$\mathcal{V}$ and $\mathcal{R}$ are defined as limits, let $\mathcal{V}_u$ and $\mathcal{R}_u$ be the expressions in Definitions 14 and 15 that converge to them respectively. Since we are mostly interested in the behavior of the system and not really in the exact values of the propagation speed and the infection rate, it is enough to set a large enough value for u, consider that $\mathcal{V} \approx \mathcal{V}_u$ and $\mathcal{R} \approx \mathcal{R}_u$, and interpret the results.

6.2 Simulation results

For all simulations, unless otherwise stated, parameters are set by default as follows: ($u = 3.5km, H = 10km, \lambda = 50km^{-2}, \theta = 3km^{-1}, v = 5km/h, \rho = 20s, r = 200m$). where H is the side length of the square surface containing the map. We assume $dt = 0.9\rho$. Each value in the diagrams we will present later is the average over 20 simulations with the same set of parameters. In the diagrams where λ does not vary, we use the same 20 maps for all the points.

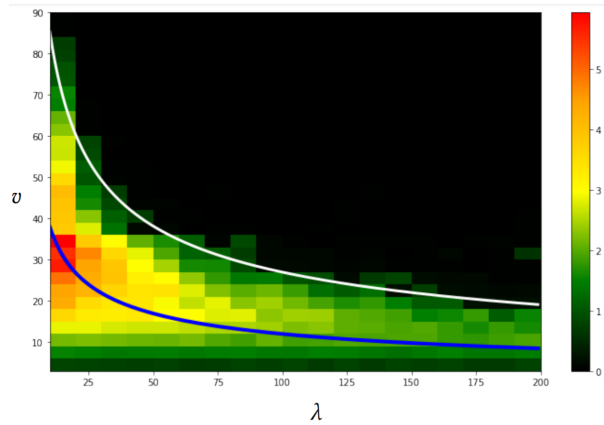

Figure 1: Infection rate $\mathcal{R}$

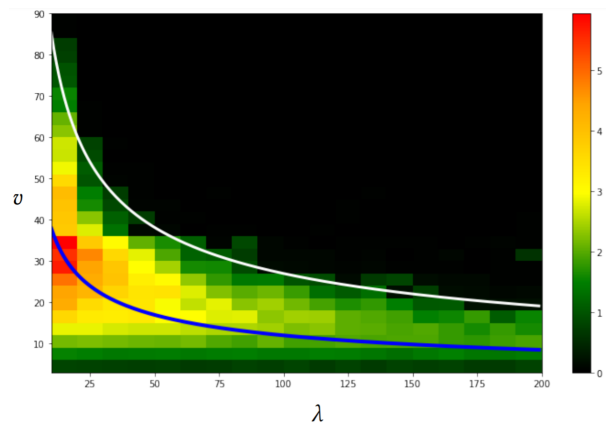

Figure 2: Propagation speed $\mathcal{V}$ in (km/h)

6.2.1 The threshold $\sqrt{\lambda}\rho v$. The critical regimes seen in Section 5 are relevant and confirm the intuitive expectations one may have for the virus propagation. However, the most remarkable result concerns the regime $\sqrt{\lambda}\rho v \gg 1$, because the lower bound found for $\mathbb{E}[\tau]$ grows with a speed of $x \mapsto \exp(x^2)/x$ in the quantity $\sqrt{\lambda}\rho v$, we can thus expect to observe a rather tight threshold at the level of which the propagation is no longer possible. To have meaningful results, we will vary λ from 10 to 200 and the speed of the agents from 1 to 90, and the other parameters will be set by default as in Sections 6.2. However, when λ is very large, the number of agents $\mathbb{E}[|\mathcal{A}|] = 2\sqrt{\lambda}H^2\theta$ will be also large since, even if it is only proportional to $\sqrt{\lambda}$, the multiplicative constant is large. To keep a reasonable number of agents, we use maps with side-length $H_\lambda := 20\lambda^{-1/4}$ for each value of λ, and the stopping propagation radius $u_\lambda := 0.45 \times H_\lambda$ to have $H > 2u$. This will guarantee that the expected number of agents is $\mathbb{E}[\mathcal{A}] = 2400$ ($\theta = 3$), and the side-lengths will vary from ≈ 11.24 to $\approx 5.32km$.

We observe in Figures 1 and 2 that the rate of infection and the speed of propagation both cancel out above a certain threshold curve, having a shape of type $v(\lambda) = c/(\rho\sqrt{\lambda})$, as indicated in blue (for $c = 2/3$) and white (for $c = 3/2$). This confirms the hypothesis of the exponential lower bound of $\mathbb{E}[\tau]$, although it is obtained with a simplified mathematical model. It seems however that this threshold is sharper for $\mathcal{R}$ than for $\mathcal{V}$. The reason why we have

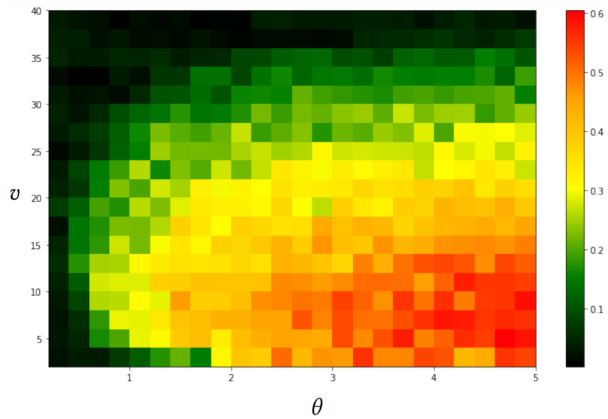

Figure 3: Infection rate $\mathcal{R}$

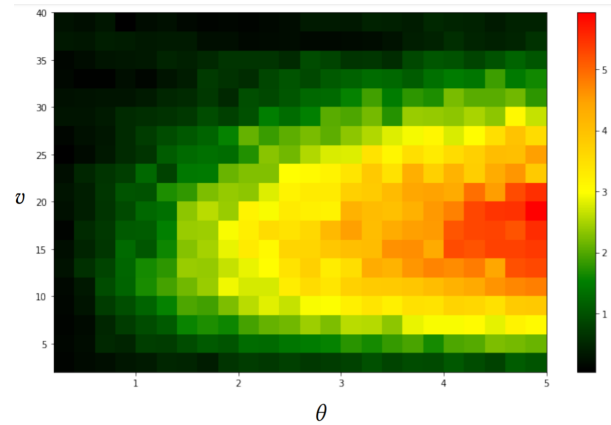

Figure 4: Propagation speed $\mathcal{V}$ in (km/h)

such a threshold is that the distribution of the edges lengths in a PVT makes it very rare to have edges much larger than the mean edge length $\mathbb{E}[L_\lambda] = 2/(3\sqrt{\lambda})$ (see Corollary 8), and therefore, when there is no edge larger than ρv, the virus cannot propagate since connection require agents to be on the same street.

With respect to Figure 2, we see that the virus can hardly propagate if $\sqrt{\lambda}\rho v \geq 3/2$. A surprising remark is that the maximum infection rate is always not far below the curve $\sqrt{\lambda}\rho v = 2/3$, while the maximum propagation speed seems to be achieved exactly at the points verifying this equation. We also observe a lower threshold value of the speed: the virus hardly propagates for $v = 3$, but as soon as $v = 6$, we see a remarkable jump in the values of $\mathcal{R}$ and $\mathcal{V}$. It is to be expected to have a weaker propagation for the small values of the speed because in the limit $v = 0$ the virus can propagate at most in the street where it was initially placed.

The third observation is that the virus propagation becomes slower as λ becomes larger. The reason is that, as predicted by the simplified model in Section 5, when $\sqrt{\lambda}$ becomes much larger than θ, we have too many streets compared to the number of agents, and therefore a_{I_0} will only meet a few agents.

6.2.2 How is the propagation speed impacted by θ and v? The propagation speed of the virus is certainly a function of all the parameters of our model. However, the distance r is given by the technology and cannot be changed, and the intensity of streets λ is known for a given city. Now, for a given malware, we want to see the influence of the intensity and speed of users on the propagation speed and the infection rate. In fact, agents that move fast enough but not too fast, i.e., not to have $\sqrt{\lambda}\rho v \geq 3/2$, will rapidly carry the virus to the other edges and facilitate its spreading. Also, when agents' intensity is important, there will be always agents on these streets that will get infected and carry the virus further.

Considering Figure 4, we see that the propagation speed and the infection rate show different behaviors. Indeed, although both are increasing in θ, $\mathcal{R}$ is maximal for v around $7 - 10km/h$, while $\mathcal{V}$ is maximal for v around $15 - 20km/h$. Moreover, the high values of the propagation speed are more concentrated while those of $\mathcal{R}$ seem to be more spread out. Also, for every θ, there is clearly an increase and then a decrease of $\mathcal{V}$ when we increase v, going from $\approx 0km/h$ to the maximal value and then returning to $0km/h$. But the value $\mathcal{R}$ does not change a lot in the first range of values of v. This means that, when agents are slow, they will stay sufficiently long on every street and therefore, once an infected agent reaches a street, it will infect many agents being on it too. Propagation speed is nevertheless slow because agents take a lot of time before exiting each street and carrying the virus to the next one. This correlation between $\mathcal{R}$ and $\mathcal{V}$ confirms the need to study them both.

Returning to the results of Figure 3, the value of v for which $\sqrt{\lambda}\rho v = 2/3$ is $v_0 \approx 16.97$. Thus, we have again that $\mathcal{R}$ is maximal in the region below the level line $\sqrt{\lambda}\rho v = 2/3$, and $\mathcal{V}$ is maximal exactly in its close neighborhood. This property would therefore be true even when varying θ. For larger values of v, we expect that the virus will not propagate anymore because the streets are not long enough, and we already see the beginning of this behavior. However, we notice that the speed at which the propagation weakens depends on θ: the higher the intensity of the agents, the higher the speed needed to weaken the virus propagation, which is to be expected since the increase of θ favors the propagation of the virus. Moreover, for small values of θ, the propagation never takes place whatever the value of v because the agents are few and do not establish enough connections (θ is below the percolation threshold).

7 CONCLUSION AND FUTURE WORK

This paper presents a novel ABM for analyzing malware propagation dynamics in D2D networks. This approach, traditionally applied for complex systems, allows us to obtain relevant and surprising findings about malware propagation in D2D, which demonstrate also the effectiveness for such dynamical communication networks. Notably, malware propagation was not possible above a first threshold ($\sqrt{\lambda}\rho v > 3/2$) and was maximal around a second threshold ($\sqrt{\lambda}\rho v = 2/3$), which corresponds to having an average length of streets equal to the distance traveled by an agent during the time ρ (needed for infection transmission). This shows the importance of street system characteristics, which has been traditionally neglected when studying malware propagation in D2D. We believe that the ABM approach has a great potential for studying malware spread in D2D communication networks. Besides generalizations such as adding attributes for the street widths, devices out of the street system or sojourn times , as future work, we aim to model and simulate countermeasure policies for reversing malware attacks.

REFERENCES

[1] A. C. Berry. 1941. The accuracy of the Gaussian approximation to the sum of independent variables. *Trans. Amer. Math. Soc.* 49 (1941), 122–136.

[2] Tony Blakely, Jason Thompson, Laxman Bablani, Patrick Andersen, Driss Ait Ouakrim, Natalie Carvalho, Patrick Abraham, Marie-Anne Boujaoude, Ameera Katar, Edifofon Akpan, et al. 2021. Determining the optimal CovidFWiopt-19 policy response using agent-based modelling linked to health and cost modelling: Case study for Victoria, Australia. *medRxiv* (2021).

[3] Kenneth A. Brakke. 2005. Statistics of Random Plane Voronoi Tessellations.

[4] M. Majid Butt, Indrakshi Dey, Merim Dzaferagic, Maria Murphy, Nicholas Kaminski, and Nicola Marchetti. 2020. Agent-Based Modeling for Distributed Decision Support in an IoT Network. *IEEE Internet of Things Journal* 7, 8 (2020), 6919–6931. https://doi.org/10.1109/JIOT.2020.2976802

[5] S.N. Chiu, D. Stoyan, W.S. Kendall, and J. Mecke. 2013. *Stochastic Geometry and Its Applications.* Wiley.

[6] Alexis Drogoul, Patrick Taillandier, Benoit Gaudou, Marc Choisy, Kevin Chapuis, Nghi Huynh Quang, Doanh Nguyen-Ngoc, Damien Philippon, Arthur Brugière, and Pierre Larmande. 2020. Designing social simulation to (seriously) support decision-making: COMOKIT, an agent-based modelling toolkit to analyse and compare the impacts of public health interventions against Covid-19. *Review of Artificial Societies and Social Simulation-RofASSS* (2020).

[7] C.-G. Esseen. 1942. On the Liapunoff limit of error in the theory of probability. *Ark. Mat. Astron.Fys.* A28, 9 (1942), 1–19.

[8] C. Gloaguen and E. Cali. 2017. Cost estimation of a fixed network deployment over an urban territory. *Annals of Telecommunications* (nov 2017). https://doi.org/10.1007/s12243-017-0614-3

[9] C. Gloaguen, F. Fleischer, H. Schmidt, and V. Schmidt. 2006. Fitting of stochastic telecommunication network models via distance measures and Monte–Carlo tests. *Telecommunication Systems* 31 (2006), 353–377.

[10] C. Gloaguen, Florian Voss, and V. Schmidt. 2009. Parametric distance distributions for fixed access network analysis and planning. 1 – 8.

[11] A. Hinsen, B. Jahnel, E. Cali, and J.-P. Wary. 2020. Malware propagation in urban D2D networks. In *2020 18th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks.* 1–9.

[12] Alexander Hinsen, Benedikt Jahnel, Elie Cali, and Jean-Philippe Wary. 2020. Phase transitions for chase-escape models on Poisson–Gilbert graphs. *Electronic Communications in Probability* 25 (2020), 1 – 14.

[13] Nicholas J. Kaminski, Maria Murphy, and Nicola Marchetti. 2016. Agent-based modeling of an IoT network. In *2016 IEEE International Symposium on Systems Engineering*. 1–7. https://doi.org/10.1109/SysEng.2016.7753151

[14] Jackie Kazil, David Masad, and Andrew Crooks. 2020. Utilizing Python for Agent-Based Modeling: The Mesa Framework. In *Social, Cultural, and Behavioral Modeling*, Robert Thomson, Halil Bisgin, Christopher Dancy, Ayaz Hyder, and Muhammad Hussain (Eds.). Springer International Publishing, Cham, 308–317.

[15] Mariusz Maziarz and Martin Zach. 2020. Agent-based modelling for Sars-CoV-2 epidemic prediction and intervention assessment: A methodological appraisal. *Journal of Evaluation in Clinical Practice* 26, 5 (2020), 1352–1360.

[16] Sebastian A Müller, Michael Balmer, William Charlton, Ricardo Ewert, Andreas Neumann, Christian Rakow, Tilmann Schlenther, and Kai Nagel. 2021. Predicting the effects of Covid-19 related interventions in urban settings by combining activity-based modelling, agent-based simulation, and mobile phone data. *medRxiv* (2021).

[17] Muaz Niazi and Amir Hussain. 2009. Agent-based tools for modeling and simulation of self-organization in peer-to-peer, ad hoc, and other complex networks. *IEEE Communications Magazine* 47, 3 (2009), 166–173. https://doi.org/10.1109/MCOM.2009.4804403

[18] Muaz A. Niazi and Amir Hussain. 2011. A Novel Agent-Based Simulation Framework for Sensing in Complex Adaptive Environments. *IEEE Sensors Journal* 11, 2 (2011), 404–412. https://doi.org/10.1109/JSEN.2010.2068044

[19] Marco Pérez-Hernández, Badraddin Alturki, and Stephan Reiff-Marganiec. 2018. FABIoT: A Flexible Agent-Based Simulation Model for IoT Environments. In *2018 IEEE International Conference on Internet of Things and IEEE Green Computing and Communications and IEEE Cyber, Physical and Social Computing and IEEE Smart Data*. 66–73. https://doi.org/10.1109/Cybermatics_2018.2018.00045

[20] Letian Zhang, Linqi Song, and Jie Xu. 2019. Preventing Malware Propagation in D2D Offloading Networks with Strategic Mobile Users. (2019), 1–6. https://doi.org/10.1109/GLOBECOM38437.2019.9014227

[21] Letian Zhang and Jie Xu. 2020. Differential Security Game in Heterogeneous Device-to-Device Offloading Network Under Epidemic Risks. *IEEE Transactions on Network Science and Engineering* 7, 3 (2020), 1852–1861. https://doi.org/10.1109/TNSE.2019.2955036