Capture the RAT: Proximity-based Attacks in 5G using the Routine Activity Theory

Ana Nieto, Antonio Acien, and Javier Lopez
Network, Information and Computer Security (NICS) Lab
Computer Science Department
University of Malaga, Spain
Email: {nieto,acien,jlm}@lcc.uma.es

Abstract—The fifth generation of cellular networks (5G) will enable different use cases where security will be more critical than ever before (e.g. autonomous vehicles and critical IoT devices). Unfortunately, the new networks are being built on the certainty that security problems cannot be solved in the short term. Far from reinventing the wheel, one of our goals is to allow security software developers to implement and test their reactive solutions for the capillary network of 5G devices. Therefore, in this paper a solution for analysing proximity-based attacks in 5G environments is modelled and tested using OMNET++. The solution, named CRAT, is able to decouple the security analysis from the hardware of the device with the aim to extend the analysis of proximity-based attacks to different use-cases in 5G. We follow a high-level approach, in which the devices can take the role of victim, offender and guardian following the principles of the routine activity theory.

Keywords—Proactive security, proximity-based cybercrime, 5G security.

I. INTRODUCTION

5G networks are now a reality which is starting to be implemented, and they are soon to be used in our everyday life, but this scenario has certain particularities that make it sensitive to attacks of different nature [11]. Our goal is to model these scenarios in realistic simulations which make it possible for us to analyse the potential problems and offer solutions to them. For example, one of the main concerns is the vulnerability of 5G networks to be taken down by rogue agents that appear to be normal users, which is considered in this paper. 5G networks rely on technologies such as mmWave, which use small interconnected relays operating in EHF (Extremely High Frequency), ranging from 110 to 300 GHz. This means that the wavelength used is really short (from 1mm to 1cm), causing the signal to be vulnerable to interferences, such as rain or buildings, and, therefore, needing additional devices for increasing the radio coverage. These devices are very sensible to jamming and other proximity-based attacks, and other open challenges that will be detailed in Section II.

Assuming the scenario in Fig. 1 will happen, there could be users devoted to restore the network stability, and they are soon to be used in our everyday life, but this scenario has certain particularities that make it sensitive to attacks of different nature [11]. Our goal is to model these scenarios in realistic simulations which make it possible for us to analyse the potential problems and offer solutions to them. For example, one of the main concerns is the vulnerability of 5G networks to be taken down by rogue agents that appear to be normal users, which is considered in this paper. 5G networks rely on technologies such as mmWave, which use small interconnected relays operating in EHF (Extremely High Frequency), ranging from 110 to 300 GHz. This means that the wavelength used is really short (from 1mm to 1cm), causing the signal to be vulnerable to interferences, such as rain or buildings, and, therefore, needing additional devices for increasing the radio coverage. These devices are very sensible to jamming and other proximity-based attacks, and other open challenges that will be detailed in Section II.

In summary, the objective of this article is to provide a
solution for analysing proximity-based attacks and evaluate the possible traceability of local attackers. Furthermore, CRAT has been implemented for OMNET++ (v.5.3) and the software is available for download in https://www.nics.uma.es/development/crat#.

The paper is structured as follows. In Section II an overview about proximity-based attacks in 5G is provided with the aim to describe the open challenges this area. In Section III the related work is described. The open challenges concerning proximity-based attacks in 5G, together with the lack of solutions to answer to these challenges properly motivates this paper. In Section IV the three-actors attack model based in the Routine Activity Theory (CRAT) is proposed. This model is implemented and validated using OMNET++ as is detailed in Section V, where some preliminary results are shown. Finally, the conclusions and future work are detailed in Section VI.

II. PROXIMITY BASED ATTACKS IN 5G

Proximity-based attacks can be defined as those attacks that can only be executed in their entirety if the attacker is close to the target. Just a few years ago the main threat actors in proximity-based attacks in cellular networks were eavesdroppers and jammers [5], considered as passive and active, respectively. To avoid these attacks there are physical layer security techniques defined, for example, to protect the relays [5]. However, in the last years more sophisticated attacks have emerged to exploit software and hardware vulnerabilities, posing new challenges in 5G environments.

Although there is a notable technological gap between 4G and 5G technologies, the resource-constrained devices in the capillary network will remain defenseless against the most complex proximity-based attacks. This is because the major improvements will be done to improve the security in the infrastructure (and the operators have the power to change the hardware and software), but end-user devices will still remain vulnerable. One of the reasons why this is so, is because the capillary network is formed by heterogeneous devices, which do not necessarily have to be updated. It is more than reasonable to assume that as long as it is necessary to maintain backward compatibility there will be open security problems.

As a consequence, in recent years the list of potential attacks on cellular networks has grown, as is highlighted in the last ENISA 5G theat landscape report [1]. If security mechanisms capable of detecting and containing attacks in these new environments are not deployed, the attacks will be propagated faster than ever taking advantage of the (powerful and efficient) 5G infrastructure, causing serious damages to both the infrastructure and the users. For example, the following are open challenges related to proximity-based attacks in 5G environments:

- Dynamic Radio Access Networks (DyRAN). DyRAN technologies will provide fast connection and handling of resources to connect the devices to the 5G infrastructure. In order to do that, new protocols are being developed. These new protocols will have potential vulnerabilities that could be exploited by an attacker with access to the environment. The density of devices and the ability of the attackers will complicate the detection and traceability of attackers, for which the collaboration of the end-user devices can be crucial.

- D2D-based attacks. Cellular networks have traditionally made use of powerful antennas in order to connect the users. However, new applications and services which will take advantage of the power of the 5G infrastructure, will require D2D communications in order to work (e.g. safe raiding and navigation). However, this will enable the propagation of attacks using these new channels of communication between the devices.

- Network Slicing Security. Network slicing is a new concept related to how the use cases (e.g. autonomous vehicle, IoT, etc.) are handled by the 5G infrastructure. Each use case will be allocated in a separate network slice in order to address the specific requirements (e.g. in terms of frequency and data rate), which will require the devices in the slice to run the services provided by the 5G infrastructure. Network slices are defined to be isolated. However, at the end, all the slices share the same network resources. Therefore, the isolation could be broken, and, if this happens, some devices can take advantage of this fact to propagate attacks to other network slices or to obtain additional resources.

- Mobile Edge Computing (MEC) Security. MEC technologies will bring closer the Cloud to the end-user. This means that many services will be executed closer to the capillary network in order to increase the performance, making able to function some critical services which require low latency rates in order to work. However, the nodes in the MEC will have less resources to detect and to analyse the attacks than the core of the 5G network. Therefore, providing proactive security mechanisms to help to these systems to detect attacks, will be key to mitigate the propagation of the attacks from the capillary network to the 5G infrastructure.

- Cross-layer attacks. The heterogeneity of the use cases and devices in these scenarios complicates the traceability of the attacks. Moreover, the countermeasures to mitigate the propagation of the attacks in the capillary network will be not feasible without taking advantage native security features already available in end-user devices.

Some solutions have been provided to fight against the proximity-based mobile malware propagation based on trustworthy entities deployed in the network [12]. Other solutions are focused on detecting attacks on services that depend on the user’s location [7]. However, although the applications implement security mechanisms, these could be overcome by a targeted attack that exploits vulnerabilities in the nodes or in the communication mechanisms with the infrastructure.

In general, the previous open challenges have a difficult short-term solution. The extremely different use cases that will be deployed using 5G (e.g. V2V, tactile internet, industry 4.0) will indeed complicate find solutions to these and other security problems.
III. RELATED WORK

The evaluation of reactive solutions against attacks is usually done using real devices or simulators. In this last case, one of the problems that can occur is that simulations are thought for very specific use cases and they do not allow easy adaptation for different types of attacks. Another limitation is that the profile of the nodes should be static. For example, OMNET++ is not designed so that nodes can have dynamic behaviours or to include security functionalities by default. Besides, 5G features are not completely available because in fact 5G networks are still being defined. Some features are being implemented but not all modules are compatible with each other, because some modify the same libraries to implement their respective functionality.

There are some works that implement the characteristics of a 5G network in the context in which this article is framed. For example, there are LTE D2D modules available for both ns3 [9], [8] and OMNET++ [3] simulators. In particular, [3] is used to implement CRAT in Section V. Although these works do not incorporate security features, they serve as a starting point to develop our solution.

In [10] the authors propose the NETA framework. This framework is very useful to model attacks in OMNET++. However, the available implementation cannot be used in the latest versions of OMNET++, and the model would still not allow us to define dynamic applications for the nodes.

In summary, although different models have been proposed to analyse the effect of attacks, or the contagion in networks of nodes [6], they are usually static models, which do not consider the interaction between the three actors on which we focus our work. In particular, in this paper the CRAT model is proposed to analyse whether proactive security mechanisms (e.g. guardian in our model) in a cellular network are really significant to stop or to mitigate proximity-based attacks. Moreover, using CRAT it is possible to analyse proximity-based attacks based on three (dynamic) profiles, applying well-establishing principles in other disciplines that are useful in this context. In addition, the proposed model can be extended to particularize the actors to different use cases in 5G scenarios.

IV. THREE-ACTORS ATTACK MODEL (CRAT)

The CRAT model is defined based on three typical actors of the routine activity theory: victim, motivated offender and (absence of a) guardian (c.f. Fig. 1). While there are works that try to finally determine how these actors evolve in cyberspace scenarios (e.g. to define cyberattacker and cybervictim profiles, or towards the definition of offender’s resources [2]), how these well-established models can help understand and predict proximity-based attacks is not currently being considered.

One of the parameters considered in the routine activity theory is the physical location of the offender at the time the offence was committed. Unlike cyberattacks (where the location of the attacker might not be relevant because the attack can be performed from any location and easily hidden), proximity-based attacks fits very well in the analysis based on said three actors and some of their physical properties. Note that proximity-based attacks offer some chance to capture the attacker (c.f. Fig. 1). Unlike cyber-attackers, a local attacker could be traceable if there are devices in the environment with the ability to identify attacks (e.g. the guardian in Fig. 1). This assertion is of vital importance in the context of this article. Other differences between cybercrime and crime when applying the routine activity theory can be found in [13], [2].

Fig. 2 shows the scope of the solution developed. The aim is to define applications able to change its behaviour at run time, to characterize the three expected behaviours, depending on the role of the actor (offender, victim or guardian). This must be independent of the node in which the applications are running. This last characteristic is fundamental since one of our objectives is to develop an extensible solution capable of integrating heterogeneous objects using similar versions of the same application to cooperate.

Therefore, the solution proposed defines the expected behaviour of the offender, the victim and the guardian, and how the role of the node changes based on the context and the messages received. This is independent on the hardware or physical characteristics which are provided/implemented in the simulators (c.f. Section III).

The following sections describe the actors of the model and their expected behaviour.

A. Actors

As described in previous sections, the CRAT model is based on three main (active) actors: victim (named Regular in the following), offender (named Eve in the following) and guardian. The name of the model derives from the relationship between these actors, and the capability to analyse the behaviour of the network using dynamic applications which implement the role of the nodes. Therefore, during the execution, a Regular node change its role if it is finally corrupted by Eve (e.g. Fig. 4).

- Regular user: These users are modeled as nodes that send packets to others users of the network with a normal frequency that is not susceptible of being regarded as suspicious.
- Eve: These nodes are used to represent malicious users who want to take down the network. As discussed before, this could be done in several ways. In this case, Eve nodes are modelled as nodes behaving as Regular nodes until a point where they decide to send Eve packets, which are targeted to the base station or to nearby nodes.
- Guardian: In contrast to Eve nodes, guardians are meant to prevent the network from shutting down when they detect some problem. Also, they could stop certain nodes from working, kicking them out of the network temporarily, if they detect suspicious activity or misbehaviour. For example, a guardian could be implemented as part of the 5G infrastructure to protect resource constrained devices in the capillary network (e.g. integrating this functionality in powerful relays).

Fig. 1 shows the aforementioned actors. Note that the base station is the door to the 5G infrastructure. The nature of this element is inherent to the 5G context, and our solution does not modify the behaviour of the base station or relays.
In other words, our solution uses the predefined behaviour of the simulation environment for the components of the infrastructure (e.g. base station).

The CRAT model can be used to analyse two type of attacks: (i) direct attacks against users by D2D communications and (ii) attacks against the services provided by the 5G infrastructure initiated by local attackers.

B. Expected behaviour - Initial assumptions

The behaviour of the actors is governed by a set of rules defined below.

**R1** A Guardian node never changes its role. It means, that we presume that a guardian cannot be infected or compromised; it is a trustworthy entity.

**R2** A Regular node can always be infected when “Eve” is present. We assume that it has no security mechanisms to detect or to react to an attack.

**R3** A Regular node compromised (or infected) by Eve can be cured (or recovered) by a guardian, but the original Eve cannot (c.f. Section IV-C).

**R4** We assume that a service can be affected by an attack from the capillary network (Eve). This is not easy to occur because the complexity of carrying out an attack of this kind. However, it can not be dismissed without further ado.

**R5** If a service is compromised, only the guardian can request its recovery.

The infrastructure has its own mechanisms to recover itself, but in our simulations the request will come from the guardian, because we want to evaluate the possible contribution to network security when trustworthy agents are used in the capillary network. Therefore, the role of a guardian is to restore regular nodes or infrastructure services if it detects incorrect system behaviour or a poor Quality of Service (QoS).

These rules can be represented graphically by means of a state machine, as shown in Fig. 3. The conditions EVE, GUA and REG mean messages sent by eve, guardian and regular, respectively. If a Regular node changes its state to “Eve”, that means that the node has been infected by some “Eve” node in the 5G network. In this scenario we focus on the contagion through D2D communications. Therefore, if a Regular node changes its state to “Eve”, that means that Eve in that moment is near said node. This can be modified if the model is adapted so that the attacks can wait for a signal (or a period) to be activated. However, the current model does not consider said behaviour, and, when the Regular node receives the message from Eve, it becomes infected.

The condition nEve \neq 0 means that the node in “Eve” mode has been infected by a node that was previously infected by another “Eve”. Therefore, we differentiate between the original Eve and an the infected Eve. This difference will be explained in more detail in Section IV-C.

Note that it is assumed that a node that is guardian never changes its state to “Eve” or “Regular”. These nodes are trustworthy entities which will be defined under particular, guaranteed security requirements (e.g. these could be special nodes deployed by the 5G infrastructure to protect from intruders itself). Indeed, if a Guardian node receives a malicious message from “Eve” trying to infect it, it will keep a report about the incident, and will be able to apply countermeasures (c.f. Section IV-D).

In addition, Fig. 5 shows the state of the services provisioning when “Eve” affects these (this last case would be very unlikely for an attacker of the characteristics indicated). In
C. Attack model

In this scenario the potential victim is the end user, even if the attacker seeks to damage the services of the infrastructure. The objective for the attacker here is to damage the end user, not necessarily representing a cost for the infrastructure, although there could be collateral damage. Moreover, as discussed above, we consider that the effect of the attack is immediate. Otherwise it could be modelled with a counter without too many changes in behaviour.

One of the objectives pursued is to use the CRAT model to identify the origin of a infection. Therefore, we differentiate between original and infected nodes. This is done under the idea that an original node (eve) will not want to be disinfected, since its goal is precisely to infect the maximum number of nodes. On the contrary, we assume that an infected node is a victim, and that, to alert the user of the infection, it will want to be disinfected.

In this article the attacker Eve follows a very simple behaviour, based on two states:

- Silent. The attacker waits a certain time before committing the attack.
- Reproduction. The attacker executes the attack against the infrastructure or a specific node (D2D).

The attacker Eve infected is a Regular node that has been infected. In this article the behaviour of the Regular node changes when it is infected to be the same as the expected behaviour for an Eve node.

D. Countermeasures

The CRAT model assumes that a “Guardian” can take countermeasures to mitigate or to stop the actions of “Eve” (Fig. 4). In this scenario, the countermeasures are taken when:

I1 The Guardian is directly attacked (c.f. Fig. 3).
I2 The Guardian detects an unjustified decrease in the quality of service (c.f. Fig. 5).

Given that it is assumed that a Regular node does not have security features, it is not possible to take into account the data provided by a regular node to take a countermeasure.

The countermeasures considered here, that could be extensible in the future, are:

C1 If I1, then the Guardian will try to cure the infected node.
C2 If I2, then the Guardian will request to restore a list of specific services.

Additional countermeasures that would be possible but that are not analysed here, are:

C3 If the Guardian detects that a Regular node just change its behaviour (I3), then the Guardian sends a message GUA Restore to the compromised node. Unlike C2, this countermeasure is based on assumptions of the Guardian, since it has not been directly attacked.
C4 If C3 is not effective (I4), then the compromised node is isolated or expelled.
C5 If I1, then C1 is applied and, in addition, the Guardian keeps a report of the incident and alert to the infrastructure.

It is important to note that for C3 to be considered, it should be assumed that the Guardian is listening proactively. In the approach followed in this article, countermeasures are defined so that the Guardian can react to direct attacks. Therefore, the current approach is less intrusive because the Guardian only reacts when an infected node (or the original Eve) attacks him. However, it is less effective to stop the propagation of this kind of attacks through the network.

Nevertheless, a very interesting future work can be the evaluation of both approaches on the security of the 5G network.
In summary, a Guardian will go through four states:

- Pre-configuration. Before its deployment in a real environment, a Guardian should be configured to understand the expected patterns in the network.
- Listening. The Guardian expects to be attacked.
- Reporting. Once Eve tries to attack the Guardian, the Guardian produces a report of the attack, saves the important artefacts, and reports the attack.
- Request restore / applying countermeasures. The Guardian responds to the attack by applying the cited countermeasures.

V. Validation of CRAT using OMNET++

The CRAT model has been implemented using OMNET++ with the SimuLTE module [3], which allows use of certain features that are not present in the out-of-the-box version, such as LTE network models or D2D (device to device) communication. This is a very important requirement in order to test our solution in an environment with characteristics of a 5G network.

The simulations developed with this tool are thought as communications between nodes, where the behaviour of each node is modelled according to the methodology detailed in Section IV. Furthermore, it has been necessary to implement additional functions to ensure that the nodes can send and receive packets at the same time.

The following sections show some preliminary results after using the model CRAT. First, the propagation attack model is detailed in order to illustrate, in an approximate way, how quickly the attacks spread in this implementation. This is fundamental to understand the behaviour of the legitimate users or Regular nodes (UE) after these are infected. Second, the life-cycle of a legitimate node (UE) is shown in order to describe how the nodes in the simulator implement the CRAT behaviour. This helps to understand the last results; the third and last step is focused on describing the role of the Guardian to mitigate the attacks. This is verified by checking that by increasing the number of guardians, the Regular nodes maintain a better behaviour for a longer period of time.

The following tests shows the effect of applying the countermeasure C1 (c.f. Section IV-D), which, in fact, is the functionality that has required more modifications to be done in the normal behaviour of UE nodes.

It is important to note that for each node with the possibility of changing the state (that is, the UE nodes), two vectors are provided to describe: (i) the number of packets sent when the node is in the state Regular (c.f. Section IV-A) , and (ii) the number of packets sent when the node is in the state Eve. This last vector corresponds to the behaviour of infected UE. Regular nodes are denoted as UE. The following figures shows the time of simulation (in seconds) on the $x$ axis, and the number of packets sent on the $y$ axis.

A. The role of Eve in the propagation attack model

In order to understand the effect of the attacker propagation model in the results shown in the following sections, it is critical to compare how the propagation speed is increased if the infected Eves can propagate the attack. Fig. 6(a) and Fig. 6(b) shows the results when only the original Eve is able to propagate the attack and when the attack can be propagated also by the infected Eves.

As a first result, note that some nodes are not infected if only the original Eve is propagating the attack. This can be
seen in the horizontal axis \( y = 0 \) in Fig. 6(a), where there are nodes that are not sending “Eve packets”, or, in other words, they do not get infected during the simulation. Instead, note that in Fig. 6(b), all the nodes end up infecting each other in less time. The results shows that, while the time for contagion is similar because the nodes are very close from the original Eve, the propagation coverage of the attack is different because, intuitively, there will be nodes outside the range of the original Eve that could only be infected through another previously infected Eve.

B. Life-cycle of a Regular node (UE)

Fig. 7 shows the life-cycle of two Regular nodes (UE0 and UE1). Note that a regular (or legitimate) node is able to change to become an infected Eve after receiving a packet from Eve (c.f. Section IV-A). For the sake of simplicity, the Guardian is not shown, but the effect of the Guardian is visible towards the changes of UE1. Besides, in Fig. 7 two behaviours are shown: on the one hand, UE0 is compromised after 2secs. and, after this becomes malicious during the rest of the simulation. On the other hand, UE1 is compromised but also recovered by the Guardian multiple times during the simulation. Note that between the 13th and 14th seconds of the simulation the node is recovered by the Guardian and continues sending normal/legitimate traffic until past the 16secs. where it is infected again.
C. The effect of the Guardian(s)

Fig. 8(a) and Fig. 8(b) shows the behaviour of the network in presence of one and two guardians respectively. The behaviour of the network is improved in the second case, given that there are two trustworthy nodes able for recover the infected Eve(s) and return these to their normal state (Regular). Therefore, it is clear that, in this case, if the number of guardians decreases then the nodes becomes malicious (Eve) faster.

Note that in Fig. 8(a), there are nodes that, once infected, never recover (e.g. UE4), or recover later (e.g. UE3). This is seen looking at y = 0. Instead, Fig. 8(a), shows a different behaviour, in which the nodes that are infected are recovered multiple times (e.g. UE3 at t = 42s and t = 52). Note that using the countermeasures defined the nodes never learn about the attack and they get infected by the same attack repeatedly. This behaviour can be improved by reserving memory for this purpose in the nodes and evaluating in turn using the model proposed in this article.

VI. CONCLUSIONS AND FUTURE WORK

In this paper the model CRAT is proposed as a solution to help in the analysis of proximity-based attacks in 5G networks. This model is based on the three actors considered in the Routine Activity Theory (victim, defender and attacker) in order to study the effect of a guardian for detecting and mitigating physical network attacks. CRAT has been implemented using OMNET++, and the results verify the relevance of the guardian to restore the network after an attack and to mitigate the effect of the attack in the capillary network. Note that the solution proposed is extensible and the behaviour of the nodes (UE, Eve, Guardian) can be enhanced in order to implement new and more complex attacks and countermeasures.

It is important to highlight that the implementation proposed is a proof of concept that for sure can be improved in the future by including other parameters and restrictions. For example, it could be very interesting to identify when (or if) the number of guardians becomes a problem for the security of the network, or even to the privacy of the surrounding nodes. Another interesting issue is to evaluate, in case the guardians are able to report the attacks, how these are integrated and correlated and/or the effect in the performance of the whole system. Finally, note that in this implementation the countermeasure only recovers the state of the infected nodes. In this case, the effect of this countermeasure does not affect (apparently) to the performance of the network. However, it will be very interesting to identify the effect of stricter countermeasures that isolate malicious nodes from the network. In these future works false positives / negatives will be a critical point to evaluate the successful of the solutions modelled using the CRAT model.

ACKNOWLEDGMENT

This work has been financed by Ministerio de Economía y Competitividad through the projects IoTest (TIN2015-72634-EXP) and SMOG (TIN2016-79095-C2-1-R). The first author has been financed by INCIBE through the grant program for excellency in advanced cybersecurity research teams.

REFERENCES


