Is Quantitative Analysis of Stuxnet Possible?

Anna Kolesnichenko\textsuperscript{1}, Pieter-Tjerk de Boer\textsuperscript{1}, Anne Remke\textsuperscript{1}, Emmanuele Zambon\textsuperscript{1}, and Boudewijn R. Haverkort\textsuperscript{1,2}

\textsuperscript{1}Centre for Telematics & Information Technology, University of Twente, Enschede, The Netherlands
\textsuperscript{2}Embedded Systems Institute, Eindhoven, The Netherlands
{\texttt{\{a.v.kolesnichenko, p.t.deboer, a.k.i.remke, emmanuele.zambon, b.r.h.m.haverkort\}@utwente.nl}

I. INTRODUCTION

Recently, many papers and reports were published on the analysis of Stuxnet’s code, [1], [2]. No quantitative analysis of Stuxnet was done yet, mostly because the necessary information is not readily available. However, quantitative analysis can be very useful, for example to obtain better insight in the spreading process and to analyze the efficiency of counter-measures. Quantitative analysis of some other computer viruses was recently done in [3], [4].

The questions we address in the present paper is to what extent quantitative analysis of Stuxnet is feasible and what is necessary to increase its precision.

II. STUXNET BASICS

Stuxnet is known as one of the most complex computer viruses that was primarily written to target Industrial Control Systems (ICSs). It was first discovered in July 2010, however Stuxnet was operating without being noticed for at least one year prior to its detection. The virus uses both known and unknown Windows vulnerabilities to install and propagate. During the propagation phase, Stuxnet behaves similarly to known worms and botnets. Once it reaches its target, it sabotages the system by reprogramming Programmable Logic Controllers (PLCs), which can lead to a disaster.

Stuxnet’s behavior consists of three phases: spreading, obtaining access to the PLC, and sabotage. In the present paper, we only address the spreading phase. Modeling of the attacking and sabotaging phases is not of interest, since once the target is reached, Stuxnet accomplished its mission. Stuxnet has the ability to propagate using many methods. We classify them for further discussion as follows (see Table 1): propagation (i) via USB flash drives and other removable media, (ii) via network, (iii) via shared folders.

Copying itself to removable drives is the main method of propagation, since ICSs are usually programmed through computers that are not connected to a network. Operators use removable drives to exchange data, and once the infected removable drive is inserted into a new computer, Stuxnet will copy itself and its supporting files. The compromised computer can infect new USB drives afterwards.

Propagation via network can be seen as a botnet or worm spread which have been recently studied and modeled, as mentioned above, [3], [4]. Note that network propagation is the only fully automatic way of spreading.

The third way of propagation includes infection via shared folders or network drives, and print spooler services. For example, Stuxnet will execute on each computer where a compromised folder is used.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>CLASSIFICATION OF PROPAGATION MECHANISMS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Local</td>
</tr>
<tr>
<td>Manual</td>
<td>Removable drives</td>
</tr>
<tr>
<td>Automatic</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

Stuxnet spreads mainly within company networks. However, propagation between networks of different companies is possible if, for example, the compromised computer has VPN connection to an outside network, or an infected USB stick is taken to the outside network (and used there).

The behavior of Stuxnet is controlled remotely. After installation, the virus contacts a command and control (C&C) server and sends information about the compromised computer. The C&C servers are mostly used for spreading new versions of the virus. However, the ability to receive information from outside can be used by attackers to help the worm propagate through specific target networks or, alternatively, stop propagation.

III. QUANTITATIVE ANALYSIS OF THE SPREADING PHASE

This section addresses the main research question: Is quantitative analysis of the Stuxnet’s spreading phase possible? Out of many approaches, in this paper we select the mean-field approach [5] since it has been shown to work efficiently and to provide fast and accurate results in similar cases [6]. The main idea of the mean-field method is to model the overall behavior of a large system via the average behavior of a single node.

Each component of the system can be modeled as a continuous-time Markov chain (CTMC). Using the published information about Stuxnet, an individual CTMC model for all relevant system components can be built, i.e., a single computer, a USB stick and a shared folder (see Figure 1a). (Note that further interaction with the security community may well lead to a refinement of this model.) Some of the transition rates are constant (e.g., removal rates (RR)), other rates depend on the states of the other CTMCs (e.g., the infection rate). Using the mean-field method the propagation model can then be built by lumping the individual CTMCs and constructing the overall CTMC for the different components. For example, the lower part of Figure 1b) depicts the accumulated mean-field model for all USB drives. The accumulated model...
of a single company network is obtained by combining all individual computers and all shared folders, describing the interaction of all the nodes within one company network. This is represented by a cloud in Figure 1b). The overall model then combines all company networks and takes into consideration the interaction between them (represented by rate $p_{ij}$) and the influence of the USB sticks (represented by rate $U_{USB}$). Note that now some rates depend on the state of all the model (e.g., rate of infection is proportional to the number of infected USB sticks, active infected computers, and infected shared folders). While the presented example only consists of three companies it is clearly possible to extend the model to more companies or even to model companies in different countries.

Once the final model is built, differential equations can be derived and solved to analyze the propagation of Stuxnet.

IV. DISCUSSION

For a model as developed in the previous section to be useful, values for many model parameters (such as infection rates) are needed. Unfortunately this is not trivial.

The automatic spreading via the network is probably the easiest to parameterize, since it does not involve humans. One could obtain values for these parameters analyzing the Stuxnet code, or possibly easier, by doing measurements on life infected computers. This is not trivial for several reasons: (i) it needs either a sufficiently large testbed, or a real target environment; (ii) accurate measurements may take a long time since Stuxnet does not tend to spread very quickly; (iii) results may be inaccurate due to the “synthetic” environment.

Aspects that involve humans are harder to parameterize; in the case of Stuxnet, this includes the propagation via shared folders and removable USB drives, and the influence of the C&C server. For some of these parameters, reasonable estimates may be made, e.g., from sociological studies. The C&C server is not included in the model. Instead, it is expected that whenever the attackers modify the virus via the C&C system, this effectively changes the parameters (and thus the behaviour) of the virus. Presumably, this does not follow a specific pattern (otherwise it could have been coded into the virus itself), so modelling this is impossible.

However, all is not bleak. Even without complete knowledge of the parameter values, and without knowledge of the influence of the C&C, potentially interesting results may be obtained. For example, by trying different values for the unknown parameters, the sensitivity of the final results to them can be ascertained, and possibly upper and lower bounds obtained. Furthermore, by comparing results from the model to data about Stuxnet’s real spreading pattern in the past, parameters can be tuned, and the actual influence of the C&C server judged.

In conclusion, we find that a mean-field model of most aspects of Stuxnet spreading could be feasible, but that parametrization is hard. This calls for a close collaboration between security and modeling experts, with interaction and benefit going both ways.