A NEW ANALYSIS OF CRYPTOLOCKER RANSOMWARE AND WELCHIA WORM PROPAGATION BEHAVIOR. SOME APPLICATIONS. III

NIKOLAY KYURKCHIEV¹, ANTON ILIEV², ASEN RAHNEV³, AND TODORKA TERZIEVA⁴

¹,²,³,⁴Faculty of Mathematics and Informatics
University of Plovdiv Paisii Hilendarski
24, Tzar Asen Str., 4000 Plovdiv, BULGARIA

ABSTRACT: In this paper we receive new models that in some situations can be applied to model computer viruses propagation. Welchia worm and Cryptolocker ransomware have a long growing phase in contrast to many other threats. In September 2013 the CryptoLocker malware starting its invasion using mainly P2P ZeuS (aka Gameover ZeuS) malware. CryptoLocker’ main aim was to receive money from the unsuspecting victims for decrypting their files. Welchia worm uses a vulnerability in the Microsoft remote procedure call service. Welchia firstly checks for Blaster worm and if it is exists continues with Blaster deletion as well as takes care for computer to be immunised for Blaster worm. Also we modeled Malicious high–risk Android App volume growth; Malware evolution; Number of users attacked by Trojan-Ransom malware; Number of users attacked by crypto-ransomware; Number of unique users attacked by Trojan-Ransom.AndroidOS.Fusob; and ”Seasonal data”. As the authors in [3] mention: “Even traffic traces used in research papers (e.g. Slammer [4] and Code-red [5]) are not public. From the published papers [4], [5] we are not able to find parameters that can be used in our model”. Many researchers make a hard efforts to describe adequately situation connected to worm propagation [15]–[63].

AMS Subject Classification: 97N50
1. PRELIMINARIES

From the perspective of applied mathematics and modeling sigmoid functions find their place in numerous areas of life and social sciences, physics and engineering, to mention a few familiar applications: population dynamics, artificial neural networks, signal and image processing antenna feeding techniques, finances and insurance.

For \( r \in \mathbb{R} \) denote by \( h_r \in \mathbb{H}(\mathbb{R}) \) the (interval) Heaviside step function given by

\[
h_r(t) = \begin{cases} 
0, & \text{if } t < r, \\
[0,1], & \text{if } t = r, \\
1, & \text{if } t > r.
\end{cases}
\]

For \( r = 0 \) we obtain the basic Heaviside step function \( h_0(t) \).

**Sums of sigmoid functions.** For a given vector \( r = (r_1, r_2, ..., r_k) \in \mathbb{R}^k \), such that \( r_1 < r_2 < ... < r_k \), and a vector \( \alpha = (\alpha_1, \alpha_2, ..., \alpha_k) \in \mathbb{R}^k \) denote

\[
H(r, \alpha; t) = \sum_{i=1}^{k} \alpha_i h_{r_i}(t).
\]

Function \( H(r, \alpha; t) \) is a step function with \( k \) steps (jumps).

Several practically important families of smooth sigmoid functions arise from population dynamics.

Sigmoid functions find multiple applications to neural networks and cell growth population models.
A classical example is the Verhulst population growth model. Verhulst model makes an extensive use of the logistic sigmoid function

\[ s_0(t) = \frac{a}{1 + e^{-kt}}. \]

**Theorem 1.** [64] For the Hausdorff distance \( d = \rho(h_0, s_0) \) between the Heaviside step function \( h_0 \) and the sigmoid Verhulst function \( s_0 \) (with \( a = 1 \)) the following inequalities hold for \( k \geq 2 \):

\[
\tilde{d}_l = \frac{\ln(k + 1)}{k + 1} - \frac{\ln \ln(k + 1)}{k + 1} \left(1 + \frac{1}{\ln(k + 1)}\right) < d < \frac{\ln(k + 1)}{k + 1} + \frac{\ln \ln(k + 1)}{k + 1} \left(1 - \frac{1}{\ln(k + 1)}\right) = \tilde{d}_r. \tag{1}
\]

The Verhulst logistic model is considered as a basic example to introduce several related mathematical problems: approximation of step and cut functions by means of logistic function, fitting a sigmoid model to time course measurement data, etc.

Similarly, one can construct sums of other suitably shifted sigmoid functions.

Here we are interested in arbitrary shifted (horizontally translated) logistic functions.

Both the step function and the logistic function preserve their form under horizontal translation—note that Verhulst equation possess constant isoclines.

Focusing on the shifted logistic function we have

\[ s_r(t) = s_0(t - r) = \frac{a}{1 + e^{-k(t-r)}}. \]

Here we consider the following superposition of sigmoid functions:

\[
S(t) = \sum_{i=1}^{n} \frac{a_i}{1 + e^{-k_i(t-r_i)}}. \tag{2}
\]

For typical example of superposition of three sigmoids see Fig. 1.

In more complicated growth situations a more precise way for describing is by our new approach by superposition of more sigmoids.

Constructive approximation theory by superposition of sigmoidal functions can be found in [67], [66].
2. ANALYSIS OF CRYPTOLOCKER RANSOMWARE INFECTION BEHAVIOR

The Cryptolocker ransomware was initiated on September 5, 2013 to May, 2014. The target of this cyberattack was OS Windows.

The malware is encrypted known types of files on the local and shared hard drives using RSA public-key cryptography.

The decryption keys are managed only by Cryptolocker servers. On the computer display window opens which contains the text for the users that they have limited time to pay using bitcoins.

There no way to control that the paying of offered amount will give back encrypted information.

Firstly we photographed the data from Fig. 2 [61].

After that we made them cumulative data:

\[
data_{\text{Cryptolocker}} := \{(0, 29032), (1, 185484), (2, 274194), (3, 309678), (4, 364517), (5, 393549), (6, 433872), (7, 493549), (8, 545162)\}
\]

We use the following model:

\[
M^*(t) = \frac{a_1}{1 + e^{-(t-0.9)}} + \frac{a_2}{1 + e^{-(t-4)}} + \frac{a_3}{1 + e^{-(t-8)}}.
\]

The fitted model for

\[
a_1 = 319354; \ a_2 = 100814; \ a_3 = 260940
\]
has the form:

$$M^*(t) = \frac{319354}{1 + e^{0.9-t}} + \frac{100814}{1 + e^{4-t}} + \frac{260940}{1 + e^{8-t}}.$$  

Here we will show how modelling approach given here will approximate these data, see Fig. 3. 

For contemporary applicable study on sigmoids and some of their applications see the monographs [9]–[14].

### 3. ANALYSIS OF WELCHIA WORM INFECTION BEHAVIOR

For epidemic as Welchia worm it is appropriately to use a model 

$$M^{**}(t) = \frac{a_1}{1 + e^{- (t-3)}} + \frac{a_2}{1 + e^{- (t-8)}} + \frac{a_3}{1 + e^{- (t-13)}} + \frac{a_4}{1 + e^{- (t-18)}} + \frac{a_5}{1 + e^{- (t-25)}},$$

for approximating data from the statistics collected on an individual Welchia [62] honeypot administered by Frederic Perriot between August 24th, 2003 and February 24th, 2004, shown in Fig. 4.
Figure 3: The fitted model $M^*(t)$.

We will explore this example by photographing the data from Fig. 4.

$$data_{Welchia} := \{\{1,1,1000\}, \{2,2333\}, \{3,3500\}, \{4,5000\}, \{5,6833\}, \{6,8000\}, \{7,9333\}, \{8,10500\}, \{9,12000\}, \{10,14000\}, \{11,16333\}, \{12,18167\}, \{13,19667\}, \{14,21000\}, \{15,22667\}, \{16,23667\}, \{17,25000\}, \{18,26333\}, \{19,27500\}, \{20,28333\}, \{21,29333\}, \{22,29500\}, \{23,29500\}, \{24,29500\}, \{25,29500\}, \{26,29500\}, \{27,29500\}, \{28,29500\}, \{29,29500\}, \{30,29500\}, \{31,29667\}, \{32,29667\}\}$$

The fitted model for

$$a_1 = 7028.35; \ a_2 = 8037.49; \ a_3 = 8426.91; \ a_4 = 5903.66; \ a_5 = 173.405$$

has the form:

$$M^{**}(t) = \frac{7028.35}{1 + e^{3-t}} + \frac{8037.49}{1 + e^{8-t}} + \frac{8426.91}{1 + e^{13-t}} + \frac{5903.66}{1 + e^{18-t}} + \frac{173.405}{1 + e^{25-t}}.$$

We receive an impressive result when approximating these data, see Fig. 5.
Figure 4: The cumulative number of Welchia attackers [62].

Figure 5: The fitted model $M^*(t)$.

4. ANOTHER APPLICATIONS

The model can be used for “seasonal data” fitting. Firstly we photographed the data from [63], see Fig. 6:
Figure 6: The "Seasonal data" [63].

\[ Seasonal\_data := \{\{0.5, 76\}, \{0.625, 72\}, \{0.75, 80\}, \{0.875, 76\}, \{1, 75\}, \{1.125, 80\}, \{1.25, 108\}, \{1.375, 105\}, \{1.5, 108\}, \{1.625, 110\}, \{1.75, 102.5\}, \{1.875, 110\}, \{2, 110\}, \{2.125, 122.5\}, \{2.25, 132.5\}, \{2.375, 146\}, \{2.5, 145\}, \{2.625, 150\}, \{2.75, 146\}, \{2.875, 141\}, \{3, 147\}, \{3.125, 158\}, \{3.25, 166\}, \{3.375, 171\}, \{3.5, 170\}\} \]

The fitted model

\[ M^{**}(t) = \frac{2874.17}{1 + e^{1-t}} - \frac{10420.4}{1 + e^{1.5-t}} + \frac{16581.7}{1 + e^{2-t}} - \frac{13286.6}{1 + e^{2.5-t}} + \frac{4580.11}{1 + e^{3-t}} \]

is visualized on Fig. 7.

5. MALICIOUS AND HIGH–RISK ANDROID APP VOLUME GROWTH

We will show how it can be modelled data in [68] for Malicious and High–Risk Android App Volume Growth, see Fig. 8.

The fitted model
A NEW ANALYSIS OF CRYPTOLOCKER RANSOMWARE...

Figure 7: The fitted model $M^{***}(t)$ for "Seasonal data".

$$M^{****}(t) = \frac{2.70186 \times 10^6}{1 + e^{1-t}} - \frac{1.17754 \times 10^7}{1 + e^{1.5-t}} + \frac{2.11802 \times 10^7}{1 + e^{2-t}} - \frac{1.85889 \times 10^7}{1 + e^{2.5-t}} + \frac{6.71282 \times 10^6}{1 + e^{3-t}}$$

is depicted on Fig. 9.

6. MALWARE EVOLUTION

We will describe how it can be modelled data in [68] for Malware Evolution, see Fig. 10.

The fitted model

$$M^{****}(t) = \frac{1207.4072532972086}{1 + e^{-0.5(-9+t)}} + \frac{860.7667574563771}{1 + e^{-0.5(-6+t)}}$$

is shown on Fig. 11.
7. NUMBER OF USERS ATTACKED BY TROJAN-RANSOM MALWARE

We will study how it can be modelled data in [69] for the number of users attacked by Trojan-Ransom malware (Q4 2014 - Q3 2015), see Fig. 12. The
A NEW ANALYSIS OF CRYPTOLOCKER RANSOMWARE... 369

Figure 10: Malware Evolution [68].

Figure 11: The fitted model \( M^{****}(t) \) for Malware Evolution.

cumulative data is:

\[
\text{Number of users attacked by Trojan - Ransom malware (Q4.2014 - Q3.2015)}_{\text{data}} := \{1, 128132\}, \{2, 278706\}, \{3, 543089\}, \{4, 880294\}\]
The fitted model

\[ M^{****}(t) = \frac{1.09329 \times 10^6}{1 + e^{-1.17649(-2.9+t)}} \]

is presented on Fig. 13.

8. NUMBER OF USERS ATTACKED BY CRYPTO-RANSOMWARE

We will explore how it can be modelled data in [70] for the number of users attacked by crypto-ransomware (November 2016 - October 2017), see Fig. 14. The cumulative data is:

\[
\text{Number of users attacked by crypto-ransomware } \text{(November 2016 – October 2017)} \text{ data :} = \{1, 192729\}, \{2, 398928\}, \{3, 497471\}, \{4, 585681\}, \{5, 660043\}, \{6, 726802\}, \{7, 824067\}, \{8, 925574\}, \{9, 980112\}, \{10, 1048504\}, \{11, 1128389\}, \{12, 1237842\}\]
The fitted model

\[ M^{******}(t) = \frac{1.39988 \times 10^6}{1 + e^{-0.5(-8.5+t)}} - \frac{1.08595 \times 10^6}{1 + e^{-0.5(-7+t)}} + \frac{1.03641 \times 10^6}{1 + e^{-0.5(-3+t)}} \]

is given on Fig. 15.

With the proposed methodology, it can also describe attacks from other viruses, for example:

- "Number of new crypto-ransomware modifications November 2016 - October 2017" [70],
- "The number of users targeted by financial malware November 2016 - October 2017" [70],
- "The number of users attacked by financial malware November 2014 - October 2015" [69],
- "Number of Trojan-Ransom encryptor modifications in Kaspersky Labs Virus Collection 2013 - 2015" [69],
- "Number of users attacked by Trojan-Ransom encryptor malware (2012 - 2015)" [69],
- "Novel ransomware attacks between 2010 and 2015" [72], [71].
9. NUMBER OF UNIQUE USERS ATTACKED BY TROJAN-RANSOM.ANDROIDOS.FUSOB

We will examine how it can be modelled data in [73] for the number of unique users attacked by Trojan-Ransom.AndroidOS.Fusob in 2016, see Fig. 16. The
cumulative data is:

\[
\text{Number of unique users attacked by } \text{Trojan - Ransom.AndroidOS.Fusob data} := \{(1, 1950), (2, 6950), (3, 10100), (4, 31100), (5, 54100), (6, 74100), (7, 86600), (8, 96600), (9, 106100), (10, 112500), (11, 117100), (12, 124100)\}
\]

The fitted model

\[
M^{\text{fit}}(t) = - \frac{178591.}{1 + e^{-0.5(-8.5+t)}} + \frac{359263.}{1 + e^{-0.4(-7+t)}} - \frac{62421.5}{1 + e^{-0.09(-3+t)}}
\]

is depicted on Fig. 17.

10. CONCLUDING REMARKS

The fitting of data to the common model (2) does not always give good results due to large number of unknown parameters: \(a_i, k_i, r_i, i = 1, 2, ..., n\).

Frivolous minimization of this functional of many variables using, for example, CAS Mathematica leads to an expected comment of the type: “...a local extremum cannot be found...”.

This indicates that the user must make the following preliminary steps:
a) careful selection of the most typical parameters for the proposed model;
b) serious data analysis and approximate determination of the parameters \( r_i \).

The construction of an interpolation polynomial for these data is not appropriate due to the high degree of polynomial (see, for example Fig. 18).

Of course, the user can use, for example, the simplified model

\[
M(t) = \frac{142.159}{1 + e^{-0.01(t-1.25)}} + \frac{144.481}{1 + e^{-1.5(t-2.25)}} - \frac{49.5313}{1 + e^{-(t-3.125)}}
\]

to approximate the “seasonal data” (see Fig. 19).

This is true in studying the cumulative number of virus attacks.

Knowing the explicit appearance of the debugging function is of paramount importance for recognizing “known viruses”.

The proposed model can be applied to model computer viruses propagation.

ACKNOWLEDGMENTS

This paper is supported by the National Scientific Program “Information and Communication Technologies for a Single Digital Market in Science, Education and Security (ICTinSES)”, financed by the Ministry of Education and Science.
REFERENCES


