Cyber Intelligence Decision Support in the Era of Big Data

Chapter · September 2015

DOI: 10.13140/RG.2.1.2981.5129

8 authors, including:

Zlatogor Minchev
Bulgarian Academy of Sciences
54 PUBLICATIONS **121** CITATIONS
SEE PROFILE

Doychin Boyadzhiev
Plovdiv University "Paisii Hilendarski"
23 PUBLICATIONS **248** CITATIONS
SEE PROFILE

Plamen Mateev
Sofia University "St. Kliment Ohridski"
42 PUBLICATIONS **291** CITATIONS
SEE PROFILE

Nina Daskalova
Sofia University "St. Kliment Ohridski"
7 PUBLICATIONS **16** CITATIONS
SEE PROFILE

Cyber Intelligence Decision Support in the Era of Big Data

Zlatogor Minchev, Georgi Dukov, Teodora Ivanova,
Kiril Mihaylov, Doychin Boyadzhiev, Plamen Mateev,
Maroussia Bojkova, Nina Daskalova

1. Problem Definition

Three key moments have to be solved for this complex problem proper approaching: (i) selection of suitable formalism for fast and easy modelling, implementing both experts' data and cyber incidents statistics on past and future cyberattacks trends; (ii) model quantification is necessary to be added, achieving a suitable machine interpretation for discrete optimization; (iii) some probabilistic elements have also to be considered, in order to achieve realistic models, practical implementation decision support, benefitting from the "big data" knowledge context of the task. Practical implementation of these moments will be given further.

2. Modelling & Results Optimization

The practical modelling for cyber incidents is organized in a simple and flexible manner, using the "*Entity – Relationship*" ("$E-R$") machine representation, successfully implemented in I-SCIP-SA environment [1] and being used in numerous cyber threats analysis [2].

The "$E-R$" representation is simplified in a graph-like form, noting the mathematical aspects of the current modelling efforts.

An oriented graph model of m nodes (representing the *Entities*) and n arcs (noting the *Relations* between entities in the model) is accomplished.

The arcs in the graph are marked in a quadratic $[m \times m]$ incident matrix $A = [a_{i,j}]$, $i = 1 \div m$, $j = 1 \div m$. The matrix A elements are binary numbers, regarding the presence ($a_{i,j} = 1$) or absence ($a_{i,j} = 0$) of an arc between the nodes i and j. For each arc $a_{i,j}$, a weighting coefficient $x_{i,j}$ are assumed.

The resulting classification of the graph nodes is calculated, using a cumulative approach for input $a_{k,j}$ arcs and their $x_{k,j}$ weights – p_k vs output $a_{j,k}$ arcs and their $x_{j,k}$ weights – q_k as follows:

$$p_k = \sum_{j=1}^{m} a_{kj}.x_{kj}, \quad q_k = \sum_{j=1}^{m} a_{jk}.x_{jk}, \qquad k = 1 \div m. \tag{1}$$

In accordance with the practical necessities of cyber incidents modelling, different R^n classification zones could be defined.

An important moment here is the difficulty, related to reverse arcs' weights recalculation. A matching necessity for initial experts' nodes classification and new future beliefs trends is expected. These decision support tasks are non-trivial, producing different optimizational complexities for the multiple objects predispositioning practical needs.

A useful quadratic approach, with Euclidean L_2 norm implementation in the classification task, is proposed in [3]. As the solutions of the quadratic optimization task are not always feasible with positive arcs' weights values, a further linear simplification with D_1 as Chebyshev (cubic) norm is accomplished.

The distance D_1 of the i^{-th} point in 2D space, following (1) the new desired predisposition $(p_i,\, q_i)$ is calculated as:

$$D_1 = \max\left(\left|\sum_j a_{i,j}.x_{i,j} - p_i\right|, \left|\sum_j a_{j,i}.x_{j,i} - q_i\right|\right), \quad i = 1 \div m,\ j = 1 \div m. \tag{2}$$

If we denote this D_1 distance with a new variable y, the following inequalities are obvious:

$$\left|\sum_j a_{i,j}.x_{i,j} - p_i\right| \le y, \tag{3}$$

$$\left|\sum_j a_{j,i}.x_{j,i} - q_i\right| \le y. \tag{4}$$

We use the fact that:

$$|x| \le a \Leftrightarrow\ -a \le x \le a. \tag{5}$$

Substituting (3) and (4) modular inequalities with a couple of linear ones, following the idea in (5), we get:

$$-y \le \sum_j a_{i,j}.x_{i,j} - p_i \le y, \tag{6}$$

$$-y \le \sum_j a_{j,i}.x_{j,i} - q_i \le y. \tag{7}$$

Thus, introducing a new nonnegative variable y, in order to achieve minimal difference from the desired new nodes cluster positioning, we have to minimize a new linear objective function Z:

$$\min Z = y, \tag{8}$$

under linear constraints:

$$\sum_j a_{i,j}.x_{i,j} \le p_i + y, \quad \sum_j a_{i,j}.x_{i,j} \ge p_i - y, \tag{9}$$

$$\sum_j a_{j,i}.x_{j,i} \le q_i + y, \quad \sum_j a_{j,i}.x_{j,i} \ge q_i - y. \tag{10}$$

3. Probability Extension

The idea behind this model extension is based on the graph arcs' existence forecasting, implementing a probabilistic approach.

For each arc $a_{i,j}$ from the matrix A, the risk r for cyber attack is defined as:

$$r_{a_{i,j}} = h_{a_{i,j}}/u_{a_{i,j}}, \quad i = 1 \div m,\ j = 1 \div m, \tag{11}$$

where h is the number of harmful requests and u – total number of requests for the $a_{i,j}$ arc.

What is also important to note here is the implemented probability distribution, benefitting from the big data knowledge context. As the combination of statistical observations, concerning past cyber incidents, have to be mixed with experts' future beliefs, a suitable approach is the Beta distribution because of its intuitive and easy implementation [4].

In this case, both the *a priori* and the *a posteriori* probabilities are defined as follows:

$$r_{a_{i,j}} \sim a\ priori\ Beta(\alpha, \beta), \tag{12}$$

$$r_{a_{i,j}} \sim a\ posteriori\ Beta((\alpha + h_{a_{i,j}}), \beta + (u_{a_{i,j}} - h_{a_{i,j}})), \quad \alpha > 0,\ \beta > 0. \tag{13}$$

4. Prototyping

The studied context is outlined, following the trends noted in [5]–[10]. The "Mobile E-trading", "Smart Mixed Realities" and "E-government Services" are selected for further exploration.

Models are created in I-SCIP-SA. The entities number is between five and seven, assuring a convenient and simple graphical illustration. 3D sensitivity diagram with four sectors (*Active*, *Passive*, *Critical* and *Buffering*) is used, following input/output (Influence/Dependence) arcs' weights cumulative assessment (see eq. (1)) with additional 3$^{\text{rd}}$ vector absolute sensitivity representation [1].

MS Excel 2010 SOLVER is used for the optimization support, due to the small nodes number and simplified interface [11]. The probabilistic cyber attacks simulation is organized in Matlab R2011b environment [12].

Three illustrative models (see Figure 1 – Figure 3) are developed, using STEMO Ltd. experts' support, working group discussions and research experience.

(i) *Botnet DDoS attack on E-government services*

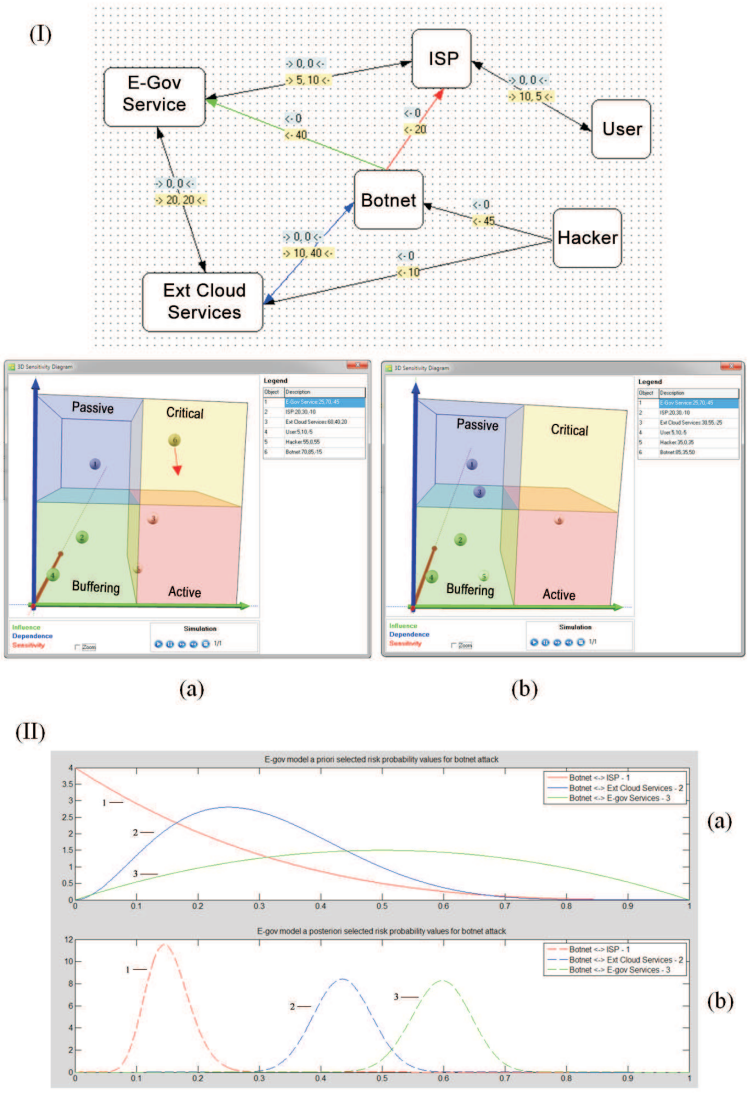

Fig. 1. E-government botnet DDoS attack system model illustration and resulting 3D sensitivity diagrams before (a) and after (b) the optimization: Panel I; Probabilistic *a priori* (a) and *a posteriori* (b) selected risk assessments: Panel II

(ii) *Bank system with credit card services compromising usage via mobile devices*

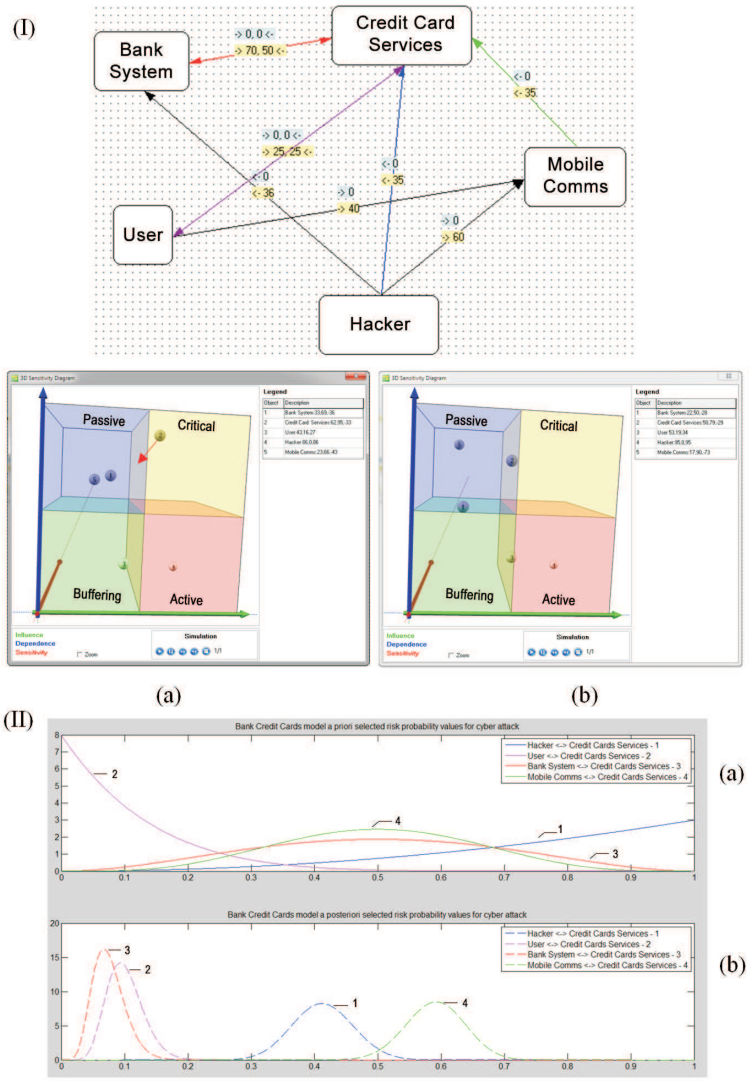

Fig. 2. Bank credit card services cyber attack system model illustration and resulting 3D sensitivity diagrams before (a) and after (b) the desired optimization: Panel I; Probabilistic *a priori* (a) and *a posteriori* (b) selected risk assessments: Panel II

(iii) *Smart home mixed reality cyber attack*

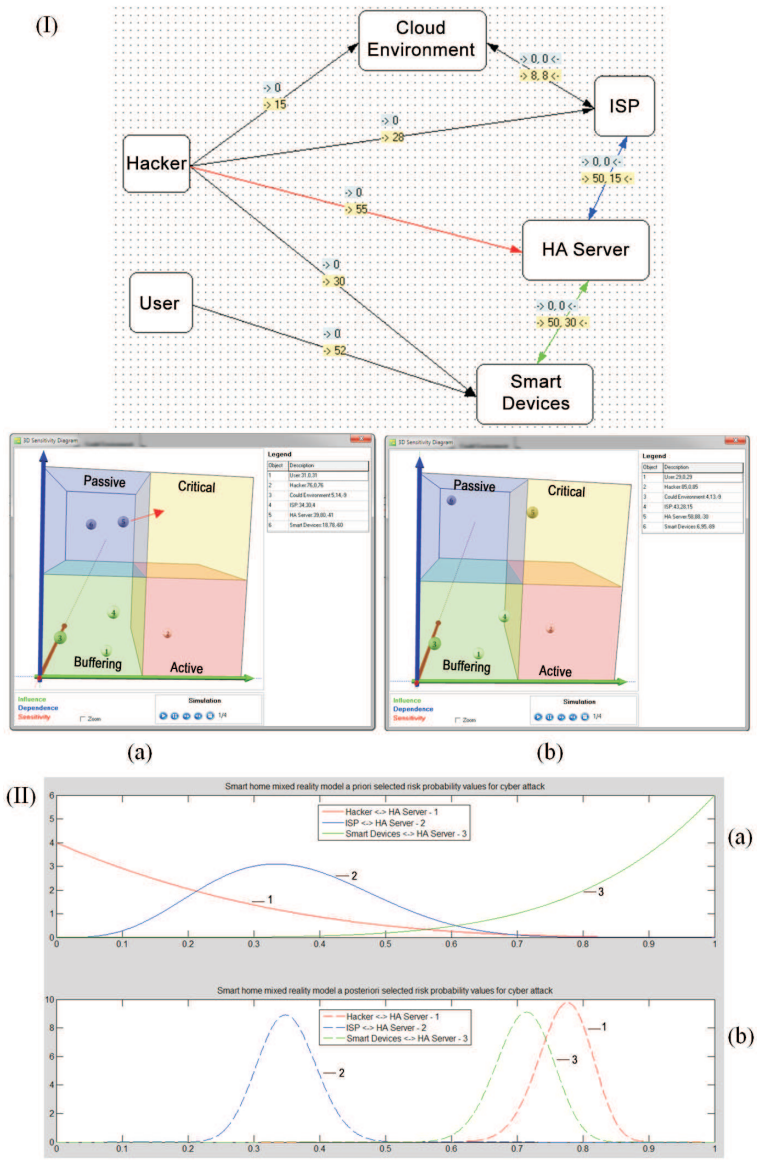

Fig. 3. Smart home mixed reality cyber attack system model illustration and resulting 3D sensitivity diagrams before (a) and after (b) the desired optimization: Panel I; Probabilistic *a priori* (a) and *a posteriori* (b) selected risk assessments: Panel II

5. Discussion

The presented complex approach for cyber intelligence decision support provides a good starting point for applied research in the era of big data.

The accomplished graph-based generic solution allows interactive analysis and near real time classification of present and future cyber threats, combining both experts' knowledge with available incidents statistics.

Additional results validation is accomplished by implementing probabilistic attacks trends evaluation, combining initial beliefs with numerical experiments simulation results.

What is also important to note here is the necessity of network sensors data integration in the presented validation approach.

This gives a possibility for deeper exploration of new cyber threats and attacks evolution, benefitting from the system modelling and assessment perspective with technologies and human factor intelligent support.

References

[1] Zlatogor Minchev & Maria Petkova, Information Processes and Threats in Social Networks: A Case Study, In Proceedings of Conjoint Scientific Seminar Modelling and Control of Information Processes, Sofia, College of Telecommunications & Post, 2010, pp. 85–93

[2] Zlatogor Minchev, Human Factor Role for Cyber Threats Resilience, In Handbook of Research on Civil Society and National Security in the Era of Cyber Warfare, Chapter 17, IGI Global, 2015, pp. 377–402

[3] Zlatogor Minchev, Numerical Optimization in Support to Graph-based Scenario Modelling, In Abstracts of 13-th Workshop on Well-posedness of Optimization Problems and Related Topics, September 12–16, Borovets, Bulgaria, 2011, p. 12

[4] Arjun Gupta & Saralees Nadarajah, Handbook of Beta Distribution and Its Applications, CRC Press, 2004

[5] Balzarotti, D. & Markatos, E. (Eds) The Red Book – A Roadmap for Systems Security Research, SysSec Consortium, 2013, `http://red-book.eu`

[6] Veselin Politov, Zlatogor Minchev, Pablo Crotti, Doychin Boyadzhiev, Maroussia Bojkova and Plamen Mateev, Cyber threats Optimization for E-

Government Services, ESGI 104 Problems & Final Reports, Sofia, September 23-27, Demetra Publishing House, 2014, pp. 77–82

[7] Threat Landscape and Good Practice Guide for Smart Home and Converged Media, Expert Report, ENISA, December, 2014, https://goo.gl/FvuLbP

[8] Zlatogor Minchev, Future Threats and Challenges in Cyberspace, CSDM Views, No. 31, Centre for Security and Defence Management, Sofia, June, 2015, http://it4sec.org/bg/system/files/views_031_0.pdf

[9] 2015 Global Megatrends in Cybersecurity, Phonemon Institute, 2015, http://www.raytheon.com/news/rtnwcm/groups/gallery/documents/content/rtn_233811.pdf

[10] 2015 Data Breach Investigation Report, Verizon, http://www.verizonenterprise.com/DBIR/2015/

[11] Christopher Zappe, Wayne Winston S. Christian Albright, Data Analysis/Optimization/Simulation Modelling with Microsoft Excel, Cengage Learning, 2011.

[12] MoonJung Cho, Wendy Martinez, Statistics in MATLAB: A Primer, Chapman and Hall/CRC Press, 2015