FUTURE CYBER ATTACKS MODELLING & FORECASTING

Zlatogor Minchev, Georgi Dukov, Doychin Boyadzhiev & Plamen Mateev

1. Problem Definition

Following leading industrial experience approaches for cyber intelligence intrusion & threats exploration (see e.g. [1], [2]), jointly with some recent research results, concerning the problem area (to note [3], [4]) three key steps have been arranged for future cyber attacks studying: (i) establishment of cyber risks landscape, implementing both intrusions and attack vectors due to expert, and literature data; (ii) supportive models for multiple intrusions risk reassessment towards expected attack vectors; (iii) experimental application of (i) & (ii) results for cyber risks landscape evolutionary prognosis. Practical realization of these steps with more details will be further outlined.

2. Cyber Risks Landscape Establishment

Solving the risk assessment is a complex target [5] that could be started with definition of rectangular matrix of intrusions vs attack vectors [$m \times n$] $R = [r_{ij}]$, $i = 1 \div m, j = 1 \div n$. The matrix R values are real numbers, regarding the risk R^* function values, calculated as follows:

$$(1)\quad R^*(T,V,I) = P(T_k|S_l) \times P(V_i|T_k,S_l) \times P\left(I_j\middle|T_k,S_l\right),$$

$$i = 1 \div m, j = 1 \div n, k = 1 \div p, l = 1 \div q.$$

Where: P (…) are the probability values, calculated for the $k^{\text{-th}}$ threat T_k vs certain $l^{\text{-th}}$ attack scenario S_l. The $i^{\text{-th}}$ vulnerability V_i and $j^{\text{-th}}$ impact I_j values are taken vs the explored T_k.

The next important exploration stage is (1) practical implementation. This could be achieved via a probabilistic multiple intrusions modelling towards landscape impacts & vulnerabilities with follow-up experimental matching to risk matrix R. The recalculation of expert beliefs is giving new opportunity for threats T evolution.

The above idea could be expressed with Dirchlet distribution [6], regarding the multidimensional case as follows:

$$(2)\quad L \sim a\ priori\ Dir\ (\alpha) \times R,$$

Where: L is the probabilistic landscape, produced as a result of expert beliefs from risk matrix R multiplication with a priori Dirichlet probabilistic cyber landscape representation – *Dir* (α), defined for k-dimensions with shape parameters α_i.

Further exploration of the problem could be achieved by intrusions modelling and cyber risk reassessment, using additional expert and literature data for better problem understanding. A suitable approach in this context is presented in the next paragraph.

3. Supportive Intrusions Modelling for Risk Reassessment

The idea for multiple cyber intrusions modelling, concerning different cyber attacks and risk matrix reassessment could be implemented, using the "*Entity – Relationship*" machine representation, combined with probabilistic forecasting, following the experience from [3], [4] but with some modifications for current tasks specifics.

The general modelling idea, after [7], is to use an oriented graph of m nodes (representing the *Entities*) and n arcs (noting the *Relations* between entities in the model).

The arcs in the graph are marked in a quadratic [$p \times p$] incident matrix $A = [a_{ij}]$, $i = 1 \div p, j = 1 \div p$. The matrix A values 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 probabilistic risk coefficient r ($0 \leq r \leq 1$) is assumed, following R^* assessment from (1).

The resulted classification of the graph nodes for the $k^{\text{-th}}$ arc is calculated, using a multiplication approach (as the studied graph model assumes to represent simultaneous events) for both input $a_{k,i}$ and output arcs $a_{k,j}$, uncertainty correcting constants $c_{k,i}$, $c_{k,j}$ (for coping model uncertainties and noisy expert data) and their $r_{k,i}$, $r_{k,j}$ risk values, producing a resulting forward (input) – R_f and backward (output) R_b probabilistic risk assessments:

$$(3)\quad R_f = \prod_{i=1}^{p} a_{k,i}.c_{k,i}.r_{k,i}, i = 1 \div p, k = 1 \div p,$$

$$(4)\quad R_b = \prod_{j=1}^{q} a_{k,j}.c_{k,j}.r_{k,j}, j = 1 \div q, k = 1 \div q.$$

The resulting model probabilistic system risk – R_s with uncertainty constant correction c_s could be defined as:

$$(5)\ R_s = c_s.(R_f / R_b).$$

Thus, in accordance with the practical necessities of cyber risk assessment a three dimensional R^3 ($x - R_f$, $y - R_b$, $z - R_s$) model probabilistic classification could be accomplished (see (1) and Section 4).

The a posteriori matrix – R' risk values of the a priori one – R (both of size $m \times n$) for the $k^{\text{-th}}$ intrusion ($k \leq m$, m – complete number of studied intrusions) are assessed with the new probabilistic risk model results R_{s_i} ($i = 1 \div p$, p – number of entities, referring to the studied risk of interest) as follows:

$$(6)\ R'_{k,j} = R_{k,j} \times \prod_{i=1}^{p} R_{s_i,j}, k \leq m, i = 1 \div p, j = 1 \div n.$$

The cyber threats landscape probabilities are reassessed after (2) for the new R'.

Finally, it is also important to note that the presented ideas have to be considered and in the dynamic sense (i.e. vs time – t) as the risk assessment is not a static process, so: $R^3 \rightarrow R^4$.

4. Experimental Application

The studied context is outlined, following the recent and future trends and prognosis, noted in [4], [8], [9]. The "Privacy & Social Engineering", "Malware & Targeted Attacks", "Data Breaching & Espionage" and "Compromised by Design Equipment" were the selected attack vectors, matched vs intrusions from: "E-mails", "Social Networks", "Web Links", "Data Sharing" and "Chat". For further exploration, following the complex nature of modern social networks, the intrusion models were aggregated around: "E-mails", "Social Networks" & "Data Sharing".

The probabilistic cyber attacks modelling was organized in Matlab R2011b environment [10].

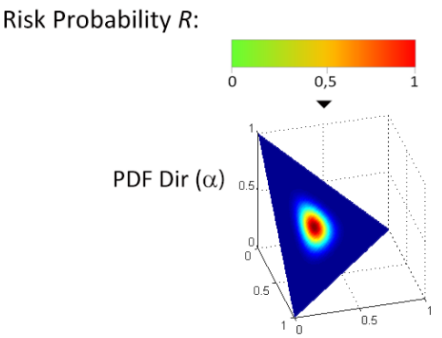
All modeling and risk probabilistic analyses were performed in I-SCIP-RA, v.1.0 software environment. The application was developed, following the "Entities" – "Relationships" ideas from I-SCIP-SA [7] and (3) – (5) for suitable probabilistic system risk assessment. Overall model entities graphical classification into a "System Risk Diagram" was organized as follows: "non-critical" & "critical", marked by the north-west/south-east main diagonal of the diagram). Additional entities system subclassification was made for "active" vs "passive" ones (denoted consecutively with white & grey colours).

Dynamically, the R^4 space of risk models and their relevant classifications of: R_f, R_b & R_s vs time – t (for a five-step time horizon up to year 2021) was given further for each attack vectors columns and intrusions rows.

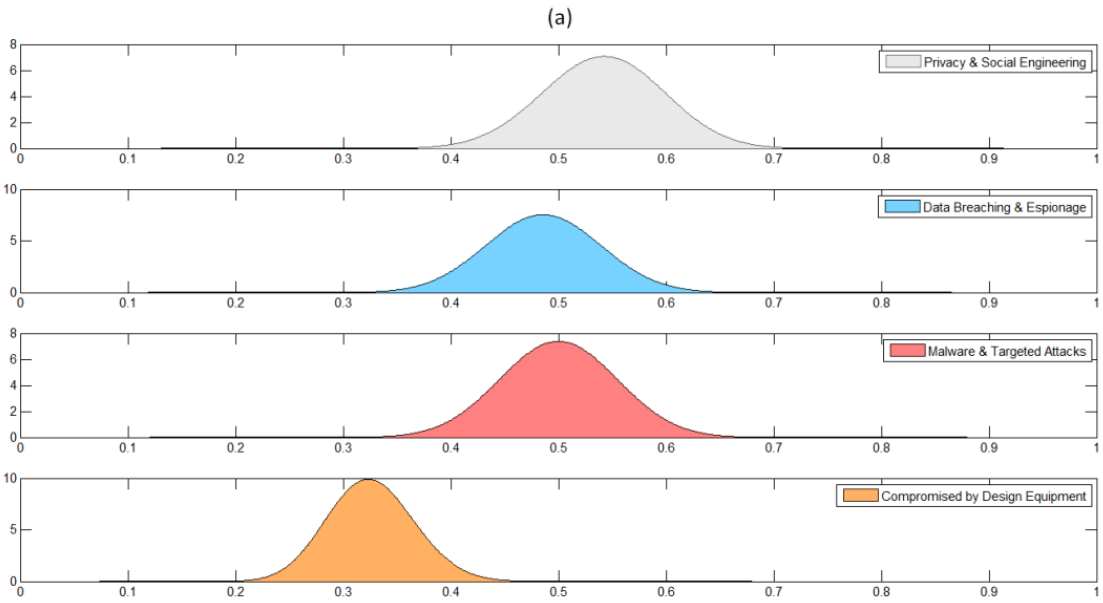
Here it should be noted that the a priori cyber landscape L probabilistic risk values and three risk system models ("E-mails", "Social Networks" & "Data Sharing" intrusions) for a posteriori risk reassessment were developed, using TechnoLogica Ltd. expert data, working group discussions and own research experience.

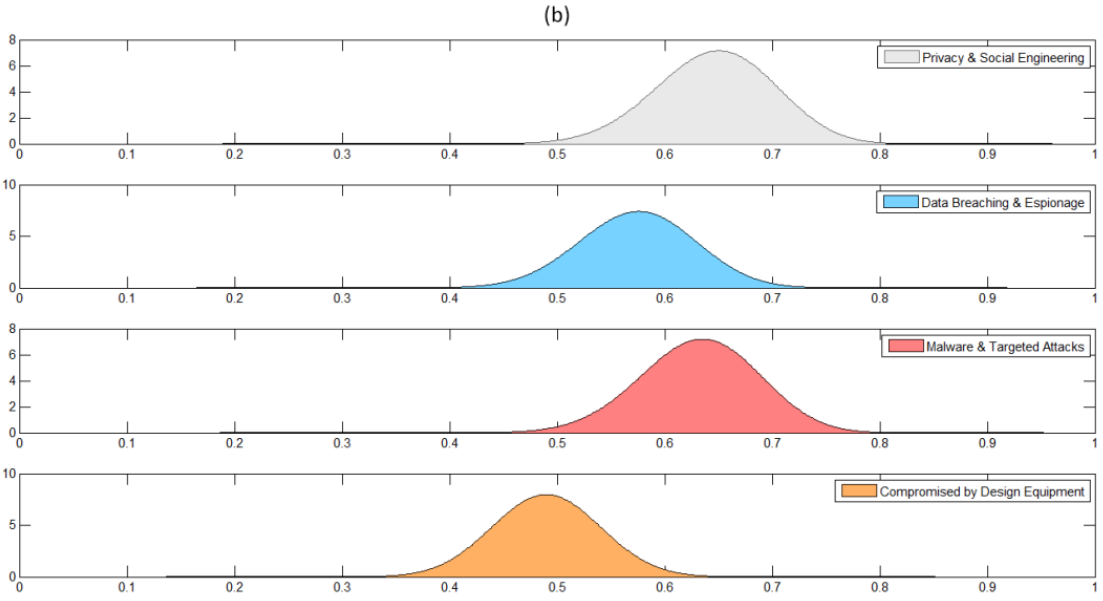
(i) Probabilistic cyber risks landscape L evolution prognosis 2016 – > 2021

Intrusions/Attack Vectors	Privacy & Social Engineering	Malware & Targeted Attacks	Data Breaching & Espionage	Compromised by Design Equipment
E-mails				
Social Networks				
Cloud Storages				

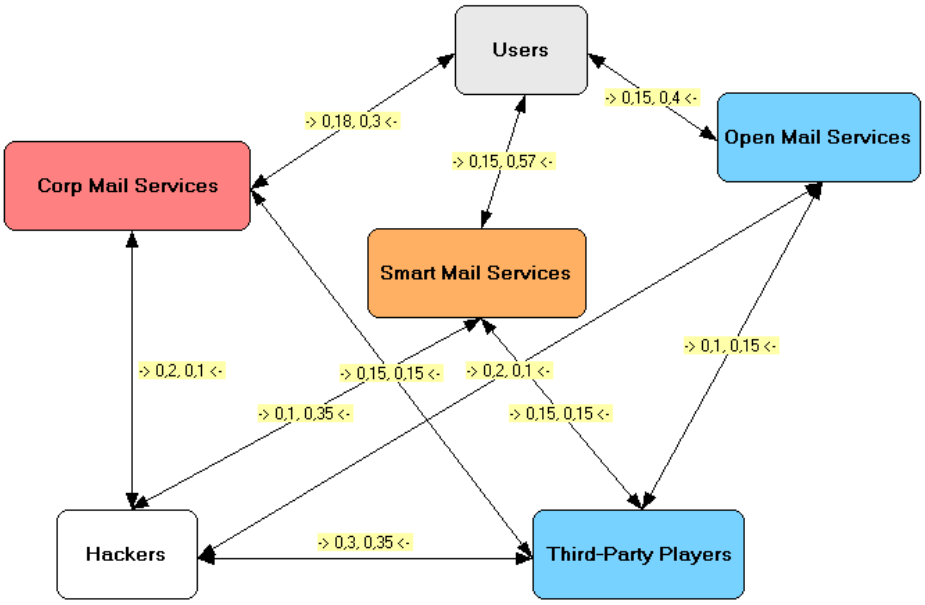


(ii) Cyber risks landscape L attack vectors a priori (a) and a posteriori (b) probabilistic distributions



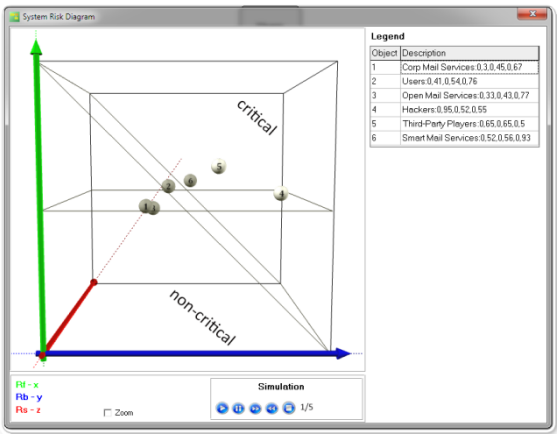


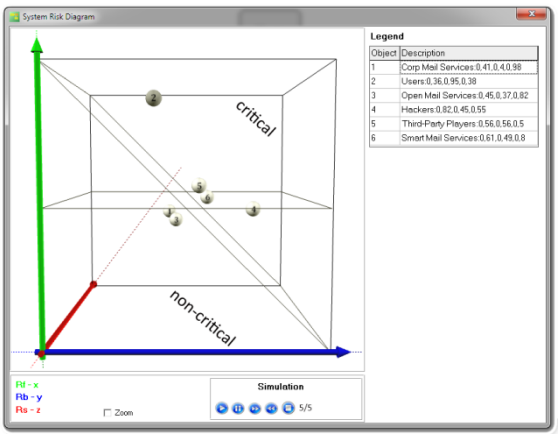
(iii.1) E-mails intrusions modelling for cyber risks landscape L reassessment



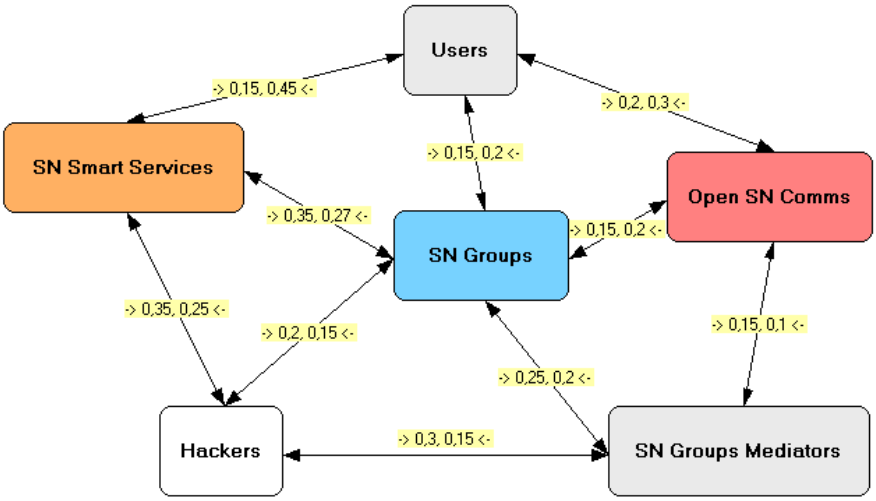
Entities vs Attack Vectors:

- Users: Privacy & Social Engineering
- Corp Mail Services: Malware & Targeted Attacks
- Third-Party Players, Open Mail Services: Data Breaching & Espionage
- Smart Mail Services: Compromised by Design Equipment



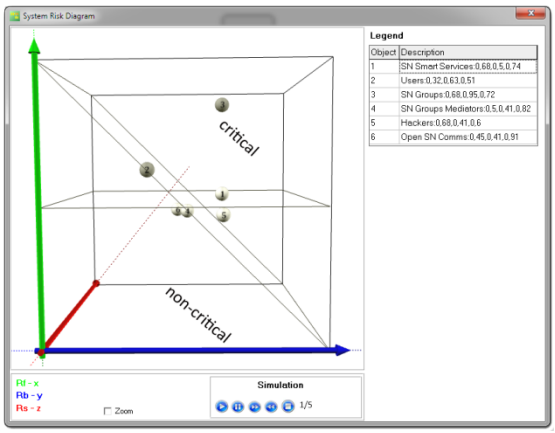


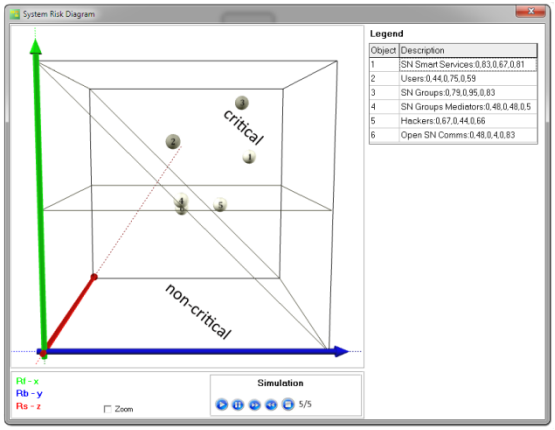
(iii.2) Social networks intrusions modelling for cyber risks landscape L reassessment



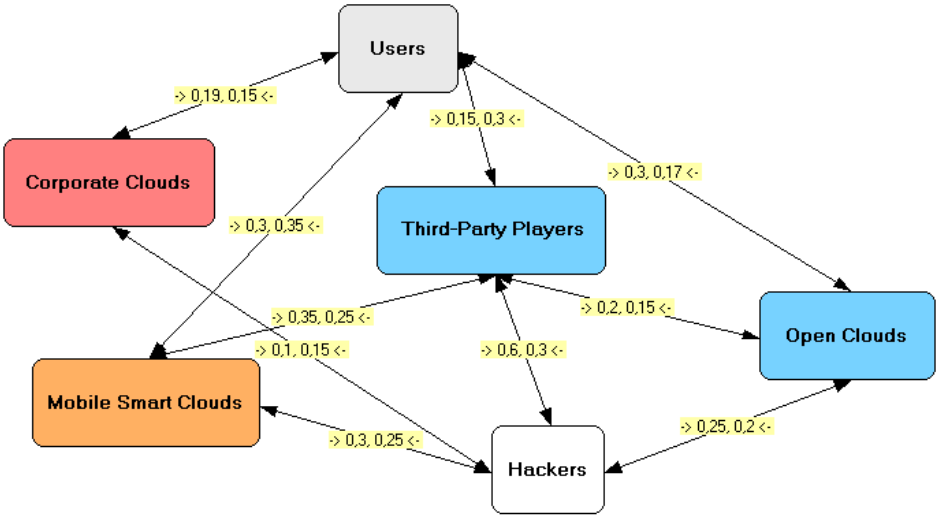
Entities vs Attack Vectors:

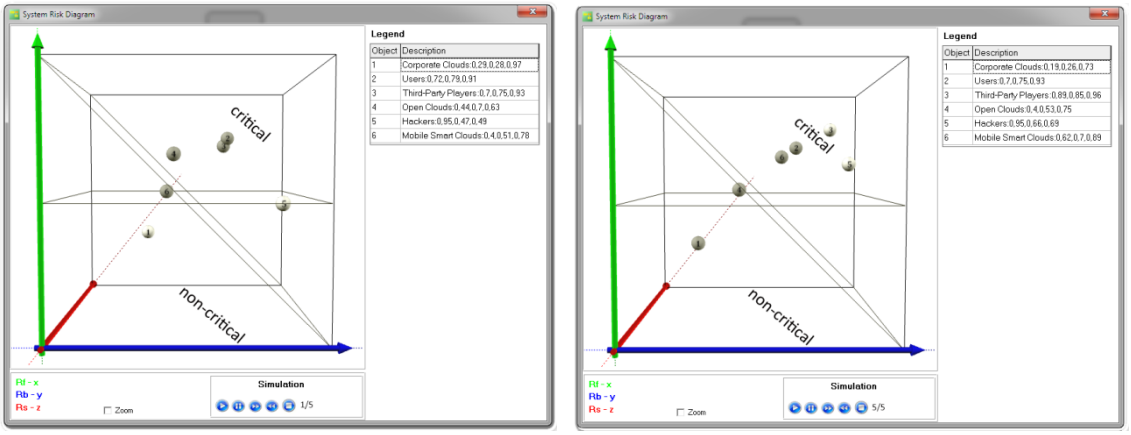
- Users, SN Groups Mediators: Privacy & Social Engineering
- Open SN Comms: Malware & Targeted Attacks
- SN Groups: Data Breaching & Espionage
- SN Smart Services: Compromised by Design Equipment





(iii.3) Cloud storages intrusions modelling for cyber risks landscape L reassessment





5. Discussion

The comprehensive understanding of future cyber attacks is producing numerous multidimensional problems that are difficult to be adequately tackled in todays' fast evolving digital world. A suitable support in this sense is proposed in the present study, combining data from both experts and literature. Further uncertainty coping is achieved with probabilistic modelling that is practically implemented in an ad-hoc designed and prototyped risk modelling and assessment research environment.

The cyber landscape risk results generalization could be summarized up to year 2021 as follows: (i) attack vectors expecting priorities of: “Privacy & Social Engineering”, “Data Breaching & Espionage”, “Malware & Targeted Attacks” with moderate ones for “Compromised by Design Equipment”; (ii) intrusions critical points for (i), encompassing: “Users”, “Open-”, “Mobile-” & “Smart-” Web 3.0 services, noting also “Third-Party Players” key role.

What however stays uncertain is the future cyber risks landscape validation and verification in advance, being generally an arguable problematic area. A useful added value in this sense could be found in the more active role of human-computer hybrid simulations, combined with detailed cyber attacks mathematical modelling.

6. Acknowledgements

The study is partially supported by SP15-FMIIT-007 project, Faculty of Mathematics & Informatics, Plovdiv University ‘Paisii Hilendarski’.

References

1. Seven Ways to Apply the Cyber Kill Chain with a Threat Intelligence Platform, Lockheed Martin, 2015, Available at: https://goo.gl/UA3Bnd

2. IT Executive Guide to Security Intelligence, IBM Security, 2015, Available at: https://goo.gl/4kPwBa

3. Minchev, Z., Dukov, G., Ivanova, T., Mihaylov, K. Boyadzhiev, D., Mateev, P., Bojkova, M., Daskalova, N. Cyber Intelligence Decision Support in the Era of Big Data, In ESGI 113 Problems & Final Reports Book, Chapter 6, FASTUMPRINT, pp. 85-92, 2015

4. Minchev, Z. & Dukov, G. Emerging Hybrid Threats Modelling & Exploration in the New Mixed Cyber-Physical Reality, BISEC 2016, Belgrade Metropolitan University, October 15, pp. 13-17, 2016

5. Rausand, M. Risk Assessment: Theory, Methods, and Applications, John Wiley & Sons, 2011

6. Ng, K. W., Tian, G-L., Tang, M-L. Dirichlet and Related Distributions: Theory, Methods and Applications, John Wiley & Sons, 2011

7. Minchev, Z. Methodological Approach for Modelling, Simulation & Assessment of Complex Discrete Systems, In Proceedings of National Informatics Conference Dedicated to the 80th Anniversary of Prof. Barnev, IMI-BAS, Sofia, November 12-13, 2015, pp. 102-110, 2016

8. Minchev, Z. Cyber Threats Identification in the Evolving Digital Reality, In Proceedings of Ninth National Conference "Education and Research in the Information Society", Plovdiv, Bulgaria, May 26-27, pp. 011-022, 2016

9. Built Environment 2050. A Report on Our Digital Future, BIM 2050 Team, 2014, Available at: https://goo.gl/iSqnm6

10. Cho, M., Martinez, W. Statistics in MATLAB: A Primer, Chapman and Hall/CRC Press, 2015