


Polymorphism and Danger Susceptibility of System Call DASTONs

View metadata, citation and similar papers at core.ac.uk

brought to you by  CORE

provided by Universiti Teknologi Malaysia Institutional Repository

Anjum Iqbal and Mond Aizami Maarof

Group on Artificial Immune Systems N Security (GAINS),
Faculty of Computer Science and Information Systems (FSKSM),
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
anjum@siswa.utm.my, maarofma@fsksm.utm.my

Abstract. We have proposed a metaphor “DANGER Susceptible daTa codON” (DASTON) in data subject to processing by Danger Theory (DT) based Artificial Immune System (DAIS). The DASTONs are data chunks or data point sets that actively take part to produce “danger”; here we abstract “danger” as required outcome. To have closer look to the metaphor, this paper furthers biological abstractions for DASTON. Susceptibility of DASTON is important parameter for generating dangerous outcome. In biology, susceptibility of a host to pathogenic activities (potentially dangerous activities) is related to polymorphism. Interestingly, results of experiments conducted for system call DASTONs are in close accordance to biological theory of polymorphism and susceptibility. This shows that computational data (system calls in this case) exhibit biological properties when processed with DT point of view.

1 Introduction

We proposed a novel metaphor [1], DANGER Susceptible daTa codON (DASTON), after having inspired from Uwe Aickelin’s proposals [2][3] and others work [3][4][5] referring Danger Theory [6][7][8][9][10] to resolve issues pertaining to self-nonsel (SNS) view point in Artificial Immune Systems (AIS). The idea of presence of DASTONs, in data processed by Danger Theory based AIS (DAIS), confers a new look towards data. The DASTONs are data chunks or various combinations of data points (data point sets) that actively participate in process for delivering required outcome. This metaphor derives its strength from important biological phenomena and substances, for example, susceptibility, host-pathogen interactions, danger signaling, codons, etc. [11][12][13][14][15][16][17][18][19].

Proposing biologically inspired metaphors for computational research involves ability to precisely map abstractions in two fields [20]. We have tried [1] to come up with analogies that help us extend our understandings and contribute more for AIS research. This paper extends the understanding of DASTON with concrete abstractions and clear experimental results.

Susceptibility might be considered a vital biological property for inferring potential danger [6][7][8][9][10] to host body and genetic polymorphism (see section 3 for details) might provide direct measure for susceptibility. Our DASTON is also highly

concerned about susceptibility. The basic research question we address in this effort is; is there any link between polymorphism and susceptibility of DASTON while studying in biological context? The answer to this question might enable us to have deeper look into metaphor and device more computational abstractions closer to biological associates.

Though, the study might be carried out for variety of data and applications, current scope is limited to system call sequences, normal and intrusion trace, available from the University of New Mexico (UNM) [21]. This data might have potential to elaborate the metaphor. Interestingly, experimental results show good compliance with theory in biology, opening new avenues for our research.

Following section 2 describes biological procedure of danger signal production by infection susceptible cell, when attacked by pathogen. Section 3 elaborates link between susceptibility and polymorphism in biology. Section 4 gives brief overview of DASTON, reader may refer [1] for details. Section 5 establishes link between polymorphism and susceptibility of DASTON in given biological context. This section portrays mythology and results of the study. Finally, section 6 concludes the effort elaborating its significance in AIS research.

2 Host Susceptibility to Pathogens: A Potential Danger

According to Polly Matzinger [6][7][8][9][10], the substances made or modified by cells under distress or suffering from abnormal death serve as danger signals for immune system. Here we introduce the term “potential danger” and link it to the infectious disease susceptibility of host. A pathogen may contribute in producing poisonous products (danger signal), leading to infectious disease, during an interplay with susceptible host. The pathogen may not interplay with unsusceptible host, hence not producing danger signal. This infers that host susceptibility might be considered as

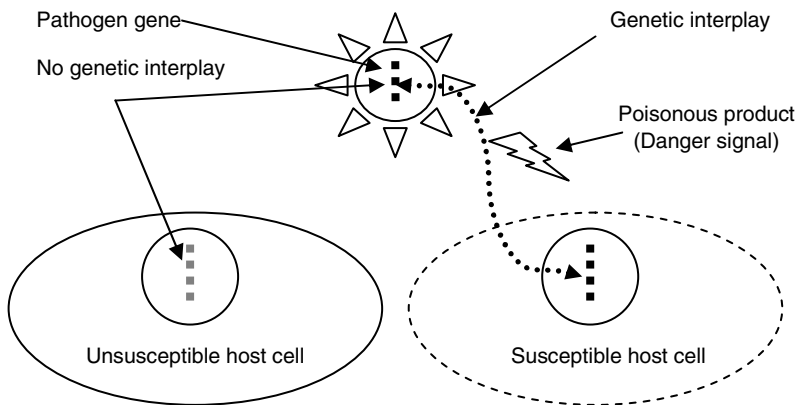


Fig. 1. The susceptible host genes are interacting with pathogen genes to produce poisonous products that cause danger signal for immune system. While, unsusceptible host genes might not interplay with pathogen. The susceptibility of host is “potential danger”.

potential danger producing factor (see figure 1). The details of infectious disease susceptibility may be better understood by reviewing related biological references [11][12][13][14][15][16][17][18][19].

The susceptibility of a host is conferred by the susceptible genetic regions. These are the high interest regions to our research. We want to closely see these regions to understand their behavior (principle) and then mapping that behavior to our case study in computation. Following paragraphs might help us understand the biological behavior of susceptible genetic regions that we will be successfully mapping to DAS-TONs in system calls data.

3 Polymorphism and Susceptibility

The word polymorphism is combination of “poly” means many and “morph” form or shape. In biology it is “the occurrence together in the same population of more than one allele (alternative form of a genetic locus) or genetic marker at the same locus with the least frequent allele or marker occurring more frequently than can be accounted for by mutation alone. Different eye colors or hair shapes result from genetic polymorphism, see figure 2.

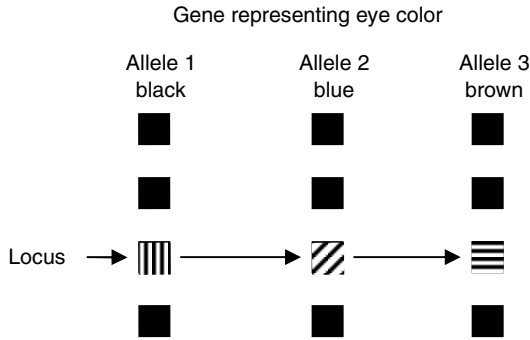


Fig. 2. A gene may have different forms (alleles) to result various phenotypes. Same gene is having alleles to result different eye colors, is an example of polymorphism.

The polymorphism may arise from single genetic unit (nucleotide) to multiple units. Currently, we are not concerned about the detailed mechanisms of biological polymorphism. We are only interested in learning that polymorphism gives rise to susceptibility. The importance of single nucleotide polymorphism (SNP) project in revealing susceptibility is worth mentioning (<http://snp.cshl.org/>). The polymorphism of tumor necrosis factor (TNF) gene is related to susceptibility of hepatitis B virus infection [22]. The major histocompatibility complex (MHC) includes the highly polymorphic human leukocyte antigen (HLA) genes that confer susceptibility to various infections including malaria, tuberculosis, HIV infections, and hepatitis B [23][25]. Polymorphism of a gene related to interleukin imparts susceptibility to hepatitis C [24] and other infections [26]. This suggests that polymorphism might be linked to potential danger susceptibility.

4 What Is DASTON?

Based on the biological concept, briefly described in section 2, we have proposed the presence of DASTONS (DANGER Susceptible daTa codON) in data [1]. These are the data chunks or combination of data points, DATONS (DATa codONS), present in data heap that actively participate in data processing to retrieve specific information from that data when subjected to triggering data or process (figure. 3). It is like presence of genetic segments in host that are susceptible to pathogenic interactions resulting in the production of toxic substances signaling danger (see figure 1). The type and size of DATONS may depend upon the nature of application, data type, and depth of details required from the data. Real examples might be that; a) only potential fields in a database might interact with query fields to result required information, and b) only potential system calls in a process might interact with exploit scripts to compromise the attacked system. One may exploit his own creative analogy to implement this biologically inspired idea. The success of analogy depends on the degree of creativity and clarity in understanding the biological concept upon which it is based [20].

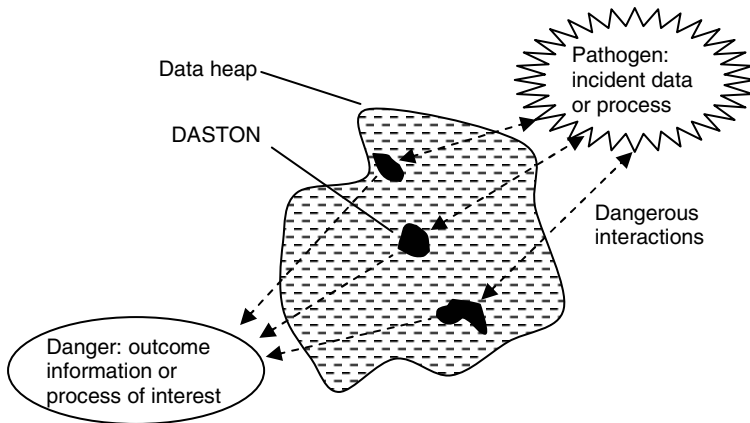


Fig. 3. DASTON present in data heap interact with incident data (named as pathogenic data) to produce required information or process (analogous to danger in danger theory)

5 Polymorphism and Susceptibility of System Call DASTONS

As discussed in section 3, biological polymorphism is “the occurrence together in the same population of more than one allele or genetic marker at the same locus”. It is also learned that biological polymorphism might provide the susceptibility measure of host body for infectious diseases [22][23][24][25][26]. We have applied the concept for establishing link between polymorphism and susceptibility of DASTON.

To enhance the worth of DASTON, we have conducted interesting experiments. Though, the metaphor might be mapped to various computational applications and data types but we stick to our constrained application - intrusion detection – and the dataset - system call sequences. These have potential to clearly illustrate the metaphor in given biological context (see Table 1).

Table 1. The abstractions corresponding to system call DASTONs

Abstractions	
Biology	Computation
Danger	Intrusion
Pathogen	Exploit script
Nucleotide	A system call
Hosts' genetic sequence	Sequence of system calls for a process/task
Triplet of nucleotides (Codon)	Set of system calls (DATON)
Susceptible Codon/segment	Danger Susceptible DATON (DASTON)

5.1 Methodology

We have conducted comparative analysis of normal and intrusion trace benchmark data, system call sequences, available from the University of New Maxico [21]. The system call pairs (DATONs), as shown in flow diagram of figure 5, are of three types; a) present in both normal and intrusion trace data, b) present in normal data only, and c) present in intrusion trace data only. The DATONs present only in intrusion trace system call sequences might be designated as the most susceptible system call pairs that are DASTONs.

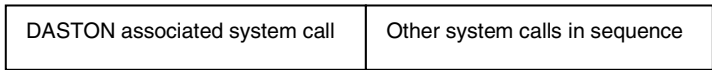


Fig. 4. Format of DASTONs associated system call pair used to get polymorphic measure of DASTONs

The “polymorphic measure” of a DASTON (system call pair present in intrusion trace sequence only) is defined as the number of distinct pairs each essentially containing one of two members from DASTON associated system calls (the system calls constituting DASTONs) , see figure 4.

In these experiments we have used the data of “synthetic sendmail” exploits (we have performed experiments with other exploits also but for simplicity presenting these results only). The normal sequences have been tested against sequences obtained from three intrusion traces (sunsendmailcp intrusion, decode intrusion, and forwarding loops). Results of experiments are in agreement with our hypothesis of “polymorphic susceptibility”, as shown in plots of figures 6 and 7.

5.2 Results

The plots of figures 6 and 7 present results with Decode Intrusion and Forwarding Loops respectively. In first experiment with Decode Intrusion there are 32 system calls associated with DASTONs, and only 3 of these have lesser polymorphic measure (number of distinct system calls combining with a DASTON associated system

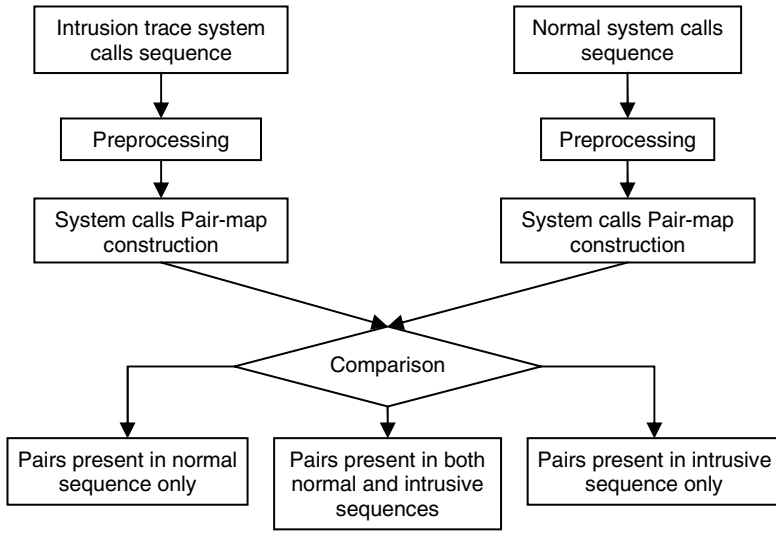


Fig. 5. Flow of the process for identifying system call DASTONs from normal and intrusion trace sequences

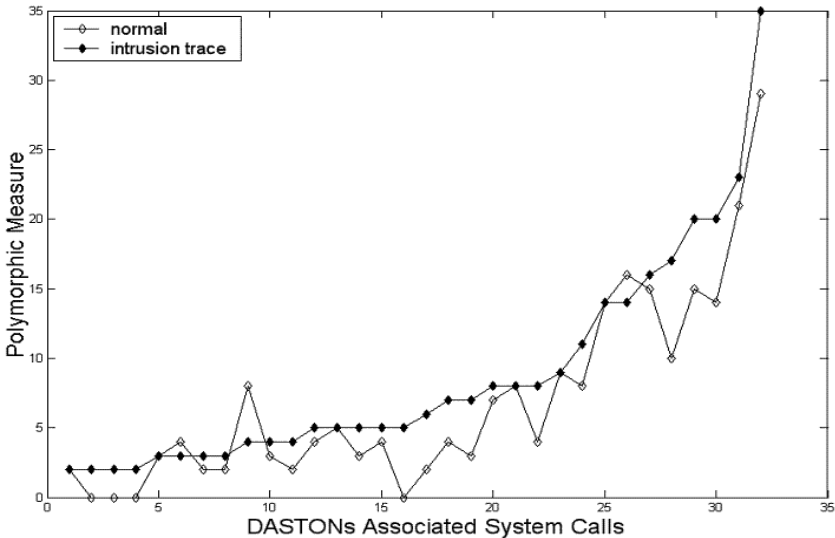


Fig. 6. Experimental results for polymorphic measure with normal and decode intrusion trace sequences

call to form distinct pairs) in intrusion trace data (filled diamond markers) than their companions in normal data (unfilled diamond marks). In second experiment with Forwarding Loops the number of DASTONs associated system calls is 35 out which only one has lesser polymorphism. This clearly demonstrates that polymorphism

would be a useful parameter to estimate Danger Susceptibility of DASTONs, like their biological buddies. It also suggests that frequency based analysis alone, of system calls sequences, should not be sufficient for describing their anomalous behavior, though it has shown success [27].

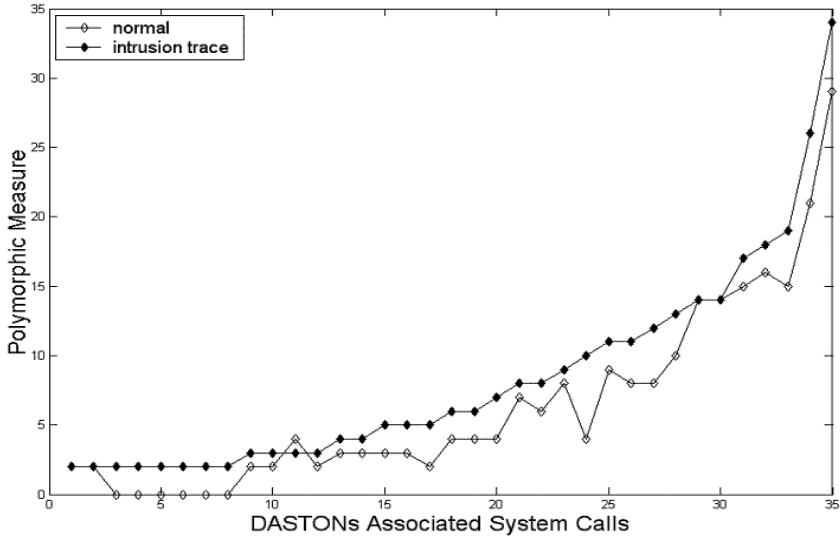


Fig. 7. Experimental results for polymorphic measure with normal and forwarding loops intrusion trace sequences

6 Conclusions

In biology, susceptibility of host cell to infectious pathogen (a case of danger producing activity) might be determined through genetic polymorphism. The same hypothesis we have applied to our proposed metaphor, DASTON, which shows compliance with the biological theory. It is a beautiful illustration of biological properties possessed by a bio-inspired computational metaphor. The data used for this illustration is system calls data that has significance in intrusion detection applications. The DASTON associated system calls have higher polymorphic values (means they combine with greater number of distinct system calls to produce distinct system call pairs) in intrusion trace sequences than those in normal sequences. Only negligible numbers of deviations appear in results. This suggests that DASTON has potential to be explored more for furthering biological abstractions in Danger Theory based AIS (DAIS) research. The idea confers a novel look towards data that DAIS processes. The established link between polymorphism and danger susceptibility recommends that frequency based analysis alone should not be sufficient for detecting anomalous behavior of system call sequences. Considering polymorphic behavior of system calls we might be able to device good anomaly detectors. Though we have successfully applied the concept to computations but immuno-informaticians and immunologists

might help to verify and further explore the immunological basis of the idea. It needs their straightforward confirmation that danger susceptibility of host for infectious pathogens is related to polymorphism of genetic segments. Their confirmation might improve the status of DASTON and bring it closer to immuno-genetics. This will open new avenues for DAIS researchers and will help devising novel computational metaphors closer to immunology theory. Remember, all this works with the creativity of the best designed machine, the human.

Acknowledgment

The authors are grateful to Ministry of Science Technology and Environment (MOSTE) Malaysia for supporting this pioneering AIS research in Malaysia. We are thankful to AIS research community for their encouraging cooperation.

References

1. Anjum Iqbal and Mohd. Aizaini Maarof (2004), Towards Danger Theory based Artificial APC Model: Novel Metaphor for Danger Susceptible Data Codons, In Proc. of International Conference on Artificial Immune Systems (ICARIS 2004).
2. Uwe Aickelin, and Steve Cayzer (2002), The Danger Theory and Its Application to Artificial Immune Systems, In Proceedings of the International Conference on Artificial Immune Systems (ICARIS, 2002), Edinburgh, UK.
3. U. Aickelin, P. Bentley, S. Cayser, J. Kim, and J. McLeod (2003), Danger Theory: The Link between AIS and IDS, In Proceedings of the International Conference on Artificial Immune Systems (ICARIS, 2003), Edinburgh, UK.
4. Emma Hart and Peter Ross (2003), Improving SOSDM: Inspirations from the Danger Theory, In Proceedings of International Conference on Artificial Immune Systems (ICARIS 2003), Springer LNCS 2787, pp. 194–203.
5. Andrew Secker, Alex A. Freitas, and Jon Timmis (2003), A Danger Theory Inspired Approach to Web Mining, In Proceedings of International Conference on Artificial Immune Systems (ICARIS 2003), Springer LNCS 2787, pp. 156–167.
6. Polly Matzinger (2002), The Danger Model: A Renewed Sense of Self, *Science*, Vol. 296, pp. 301-305.
7. P. Matzinger (2001), The Danger Model In Its Historical Context, *Scand. J. Immunol*, Vol. 54, pp. 4-9.
8. Stefania Gallucci, Martijn Lolkema, and Polly Matzinger (1999), Natural Adjuvants: Endogenous Activators of Dendritic Cells, *Nature Medicine*, Vol. 5, No. 11, pp. 1249-1255
9. Polly Matzinger, The Real Function of The Immune System, Last accessed on 06-04-04, URL:<http://cmmg.biosci.wayne.edu/asg/polly.html>
10. Polly Matzinger (1998), An Innate sense of danger, *Seminars in Immunology*, Vol. 10, pp. 399-415.
11. Michael A. Lutz, Francine Gervais, Alan Bernstein, Arthur L. Hattel, and Pamela H. Correll (2002), STK Receptor Tyrosine Kinase Regulates Susceptibility to Infection with *Listeria Monocytogenes*, *Infection and Immunity*, Vol. 70, No. 1, p. 416–418.
12. S Roy, A V S Hill, K Knox, D Griffithsand, D Crook (2002), Association of Common Genetic Variant with Susceptibility to Invasive Pneumococcal Disease, *BMJ* Volume 324, page 1369.

13. Wilfred Goldmann (2003), The Significance of Genetic Control in TSEs, *Microbiology-Today*, Vol. 30/Nov. 03, pp. 170-171
14. Jennie Blackwell (2002), Genetics and Genomics in Infectious Disease, CIMR Research Report, Last accessed on 06-04-04, URL:http://www.cimr.cam.ac.uk/resreports/report2002/pdf/blackwell_low.pdf
15. Paul M. Coussens, Brian Tooker, William Nobis, and Matthew J. Coussens (2001), Genetics and Genomics of Susceptibility to Mycobacterial Infections in Cattle, On-line publication on the 2001 IAAFSC web site.
16. Adrian VS Hill (1999), Genetics and Genomics of Infectious Disease Susceptibility, *British Medical Bulletin*, Vol. 55, No. 2, pp. 401-413.
17. Sean V. Tavtigian¹, et al (2001), A Candidate Prostate Cancer Susceptibility Gene at Chromosome 17p, *Nature Genetics*, Volume 27, pp. 172-180.
18. Jean-Laurent Casanova (2001), Mendelian Susceptibility to Mycobacterial Infection in Man, *Swiss Med Weekly*, Vol. 131, pp. 445-454.
19. P. Denny, E. Hopes, N. Gingles, K. W. Broman, W. McPheat, J. Morten, J. Alexander, P. W. Andrew, and S. D.M. Brown (2003), A major Locus Conferring Susceptibility to Infection by *Streptococcus Pneumoniae* in Mice, *Mammalian Genome*, Springer, Volume 14, pp. 448-453.
20. S. Forrest, J. Balthrop, M. Glickman and D. Ackley (2002). "Computation in the Wild." In the Internet as a Large-Complex System, edited by K. Park and W. Willins: Oxford University Press. July 18, 2002.
21. Intrusion Detection Data Sets, URL:<http://www.cs.unm.edu/~immsec/systemcalls.htm>, Last cited on 01-05-2005.
22. Yue-Su Zhou, Fu-Sheng Wang, Ming-Xu Liu, Lei Jin, Wei-Guo Hong (2005), Relationship between susceptibility of hepatitis B virus and gene polymorphism of tumor necrosis factor- α , *World Chin J Digestol*, Vol. 13, No. 2, pp. 207-210.
23. Dominic Kwiatkowski (2000), Susceptibility to infection, *BMJ*, Vol. 321, pp. 1061-1065.
24. Susanne Knapp and Branwen J. W. Hennig (2003), Interleukin-10 promoter polymorphisms and the outcome of hepatitis C virus infection, *Immunogenetics*, Vol. 55, pp. 362-369.
25. Adrian VS Hill (1999), Genetics and genomics of infectious disease susceptibility, *British Medical Bulletin*, Vol. 55, No. 2, pp. 401-413.
26. Amir H. Sabouri, and Mineki Saito (2004), Polymorphism in the Interleukin-10 Promoter Affects Both Provirus Load and the Risk of Human T Lymphotropic Virus Type I-Associated Myelopathy/Tropical Spastic Paraparesis, *Journal of Infectious Diseases*, Vol. 190, pp. 1279-1285.
27. D-K. Kang, D. Fuller, and V. Honavar, "Learning Classifiers for Misuse and Anomaly Detection Using a Bag of System Calls Representation," Technical Report ISU-CS-TR 05-06, Computer Science Department, Iowa State University, Ames, IA, USA, Mar 3, 2005.