Biologically Motivated Approaches for Complex Problem Solving

Sushil Kumar, Praneet Saurabh and Bhupendra Verma

Abstract Danger Theory is presented with particular predominance on analogies in the Artificial Immune Systems world. Artificial Immune System (AIS) is relatively naive paradigm for intelligent computations. The inspiration for AIS is derived from natural Immune System (IS). The idea is that the artificial cells release signals describing their status, e.g., safe signals and danger signals. The various artificial cells use the signals in order to adapt their behavior. This new theory suggests that the immune system reacts to threats based on the correlation of various (danger) signals and it provides a method of 'grounding' the immune response, i.e., linking it directly to the attacker. In this paper, we look at Danger Theory from the perspective of AIS practitioners and an overview of the Danger Theory is presented with particular emphasis on analogies in the Artificial Immune Systems world.

Keywords Artificial immune system · Danger theory · System cells

Introduction

Over the last decade, a new theory has become popular amongst immunologists. It is called the Danger Theory, and its chief advocate is Matzinger [1–3]. A variety of contextual clues may be essential for a meaningful danger signal, and

S. Kumar (✉) · P. Saurabh · B. Verma
Department of CSE, TIT, Bhopal, India
e-mail: asktosushil@gmail.com

P. Saurabh
e-mail: praneetsaurabh@gmail.com

B. Verma
e-mail: bk_verma3@gmail.com

V. Kumar and M. Bhatele (eds.), *Proceedings of All India Seminar on Biomedical Engineering 2012 (AISOBE 2012)*, Lecture Notes in Bioengineering,
DOI: 10.1007/978-81-322-0970-6_23, © Springer India 2013

immunological studies will provide a framework of ideas as to how ‘danger’ is assessed in the HIS. The danger signals should show up after limited attack to minimize damage and therefore have to be quickly and automatically measurable. Once the danger signal has been transmitted, the artificial immune systems (AIS) should react to those artificial antigens that are ‘near’ the emitter of the danger signal. A number of advantages are claimed for this theory; not least that it provides a method of ‘grounding’ the immune response. The theory is not complete, and there are some doubts about how much it actually changes behavior and or structure. Nevertheless, the theory contains enough potentially interesting ideas to make it worth assessing its relevance to AIS. Few other AIS practitioners are aware of the Danger Theory, notable exceptions being Burgess [4] and Willamson [5]. Hence, this deals directly with the Danger Theory. In the next section, we provide an overview of the Danger Theory, pointing out, where appropriate, some analogies in current AIS models. We then discuss about anomaly detection for AIS.

Danger Theory

The AIS are computational systems designed on the principles of natural Immune System (IS), which is highly distributed, adaptive and diverse system [6]. Danger Theory is presented with particular emphasis on analogies in the Artificial Immune Systems world [7, 8]. The idea is that the artificial cells release signals describing their status, e.g., safe signals and danger signals. The immune system is commonly thought to work at three levels: External barriers (skin, mucus), innate immunity, and the acquired or adaptive immune system. As part of the third and most complex level, B Lymphocytes secrete specific antibodies that recognize and react to stimuli. It is this pattern matching between antibodies and antigens that lie at the heart of most Artificial Immune System implementations. Another type of cell, the T (killer) lymphocyte, is also important in different types of immune reactions. Although not usually present in AIS models, the behavior of this cell is implicated in the Danger model and so it is included here. From the AIS practitioner’s point of view, the T killer cells match stimuli in much the same way as antibodies do. It is fundamental that only the ‘correct’ cells are matched as otherwise this could lead to a self-destructive autoimmune reaction. Classical immunology [9] stipulates that an immune response is triggered when the body encounters something non-self or foreign. It is not yet fully understood how this self-non-self discrimination is achieved, but many immunologists believe that the difference between them is learnt early in life. In particular it is thought that the maturation process plays an important role to achieve self-tolerance by eliminating those T and B cells that react to self. In addition, a ‘confirmation’ signal is required; that is, for either B cell or T (killer) cell activation, a T (helper) lymphocyte must also be activated. Matzinger’s Danger Theory debates this point of view (for a good introduction, see Matzinger [1]). Technical overviews can be found in Matzinger [2] and Matzinger [3]. Danger Theory clearly has many facets and intricacies, and we have touched on only a few.

It might be instructive to list a number of considerations for an Artificial Immune System practitioner regarding the suitability of the danger model for their application. The basic consideration is whether negative selection is important. If so, then these points may be relevant:

Negative selection is bound to be imperfect, and therefore auto-reactions (false positives) are inevitable.
The self-/non-self boundary is blurred since self- and non-self antigens often share common regions.
Self changes over time. Therefore, one can expect problems with memory cells, which later turn out to be inaccurate or even auto-reactive.

If these points are sufficient to make a practitioner consider incorporating the Danger Theory into their model, then the following considerations may be instructive:

1. A danger model requires an antigen-presenting cell, which can present an appropriate danger signal.
2. 'Danger' is an emotive term. The signal may have nothing to do with danger
3. The appropriate danger signal can be positive (presence of signal) or negative (absence).
4. The danger zone in biology is spatial. In Artificial Immune System applications, some other measure of proximity (for instance temporal) may be used.
5. If there is an analogue of an immune response, it should not lead to further danger signals. In biology, killer cells cause a normal cell death, not danger.
6. Matzinger proposes priming killer cells via antigen presenting cells for greater effect. Depending on the immune system used (it only makes sense for spatially distributed models) this proposal may be relevant.
7. There are a variety of considerations that are less directly related to the danger model. For example, migration—how many antibodies receive signal one/two from a given antigen-presenting cell? In addition, the Danger Theory relies on concentrations, i.e., continuous not binary matching.

This theory is borne out of the observation that there is no need to attack everything that is foreign, something that seems to be supported by the counter examples above. In this theory, danger is measured by damage to cells indicated by distress signals that are sent out when cells die an unnatural death (cell stress or lytic cell death, as opposed to programmed cell death, or apoptosis). Figure 1 depicts how we might picture an immune response according to the Danger Theory. A cell that is in distress sends out an alarm signal, whereupon antigens in the neighborhood are captured by antigen-presenting cells such as macrophages, which then travel to the local lymph node and present the antigens to lymphocytes. Essentially, the danger signal establishes a danger zone around itself. Thus, B cells producing antibodies that match antigens within the danger zone get stimulated and undergo the clonal expansion process. Those that do not match or are too far away do not get stimulated.

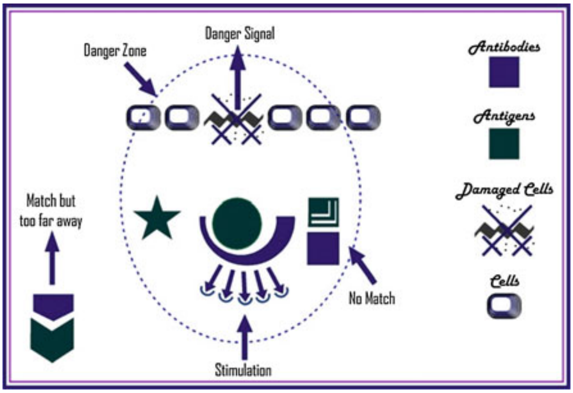

Fig. 1 Danger theory model

Another way of looking at the danger model is to see it as an extension of the Two-signal model by Bretscher and Cohn [10]. In this model, the two signals are antigen recognition (signal one) and co-stimulation (signal two).

Co-stimulation is a signal that means "this antigen really is foreign" or, in the Danger Theory, "this antigen really is dangerous". How the signal arises will be explained later. The Danger Theory then operates by applying three laws to lymphocyte behavior (the laws of lymphotics [11]):

Law 1. Become activated if you receive signals one and two together. Die if you receive signal one in the absence of signal two. Ignore signal two without signal one.

Law 2. Accept signal two from antigen-presenting cells only (or, for B cells, from T helper cells). B cells can act as antigen-presenting cells only for experienced (memory) T cells. Note that signal one can come from any cells, not just antigen-presenting cells.

Law 3. After activation (activated cells do not need signal two) revert to resting state after a short time.

For the mature lymphocyte, (whether virgin or experienced) these rules are adhered to. However, there are two exceptions in the lymphocyte lifecycle. First, immature cells are unable to accept signal two from any source. This enables an initial negative selection screening to occur. Second, activated (effector) cells respond only to signal one (ignoring signal two), but revert to the resting state shortly afterwards.

The Danger Theory and Some Affinity to AIS

Danger Theory clearly has many features and dilemmas, and we have touched on only a few. It might be instructive to list a number of considerations for an Artificial Immune System practitioner regarding the suitability of the danger model for their application. The basic consideration is whether negative selection is important. If so, then these points may be relevant:

- The self-/non-self boundary is blurred since self- and non-self antigens often share common regions.
- Self changes over time. Therefore, one can expect problems with memory cells, which later turn out to be inaccurate or even auto reactive.
- Negative selection is bound to be imperfect, and therefore auto reactions (false positives) are inevitable.

If these points are sufficient to make a practitioner consider incorporating the Danger Theory into their model, then the following considerations may be instructive:

1. A danger model requires an antigen-presenting cell, which can present an appropriate danger signal.
2. 'Danger' is an emotive term. The signal may have nothing to do with danger.
3. The appropriate danger signal can be positive (presence of signal) or negative (absence).
4. The danger zone in biology is spatial. In AIS applications, some other measure of proximity (for instance temporal) may be used.
5. If there is an analogue of an immune response, it should not lead to further danger signals. In biology, killer cells cause a normal cell death, not danger.
6. Matzinger proposes priming killer cells via antigen presenting cells for greater effect.

Depending on the immune system used (it only makes sense for spatially distributed models) this proposal may be relevant.

The Danger Theory and Anomaly Detection

In anomaly detection we watch not for a known intrusion—a signal—but rather for abnormalities in the traffic; we assume that something abnormal is probably suspicious. The construction of such a detector starts by forming an opinion on what constitutes normal for the observed subject (which can be a computer system, a particular user etc.), and then deciding on what percentage of the activity to flag as abnormal and how to make this particular decision (Fig. 2). This detection principle flags behavior that is unlikely to originate from the normal process, without needing actual intrusion scenarios [12].

In this section we will present indicative examples of such artificial systems, explain their current shortcomings, and show how the Danger Theory might help overcome some of these.

One of the first such approaches is presented by Forrest et al. [13] and extended by Hofmeyr and Forrest [15]. This work is concerned with building an AIS that is able to detect non-self in the area of network security where non-self is defined as

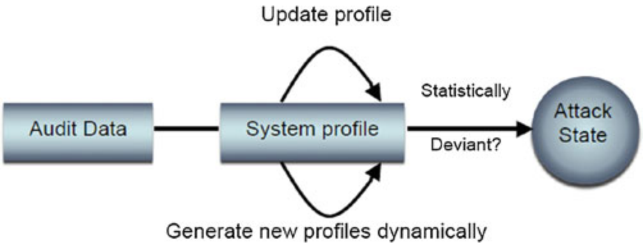

Fig. 2 A typical anomaly detection system

an undesired connection. All connections are modeled as binary strings and there is a set of known good and bad connections, which is used to train and evaluate the algorithm. To build the AIS, random binary strings are created called detectors. These detectors then undergo a maturation phase where they are presented with good, i.e., self, connections. If they match any of these they are eliminated otherwise they become mature, but not activated. If during their further lifetime these mature detectors match anything else, exceeding a certain threshold value, they become activated. This is then reported to a human operator who decides whether there is a true anomaly. If so the detectors are promoted to memory detectors with an indefinite life span and minimum activation threshold. Thus, this is similar to the secondary response in the natural immune system, for instance after immunization.

An approach such as the above is known in AIS as negative selection as only those detectors (antibodies) that do not match live on. It is thought that T cells mature in similar fashion in the thymus such that only those survive and mature that does not match any self cells after a certain amount of time.

An alternative approach to negative selection is that of positive selection as used for instance by Forrest et al. [14] and by Somayaji and Forrest [16]. These systems are a reversal of the negative selection algorithm described above with the difference that detectors for self are evolved. From a performance point of view there are advantages and disadvantages for both methods. A suspect non-self string would have to be compared with all self-detectors to establish that it is non-self, whilst with negative selection the first matching detector would stop the comparison. On the other hand, for a self-string this is reversed giving positive selection the upper hand. Thus, performance depends on the self to non-self ratio, which should generally favor positive selection.

However, there is another difference between the two approaches: the nature of false alarms. With negative selection inadequate detectors will result in false negatives (missed intrusions) whilst with positive selection there will be false positives (false alarms). The preference between the two in this case is likely to be problem specific.

What could such danger signals be? They should show up after limited infection to minimize damage and hence have to be quickly and automatically measurable. Suitable signals could include:

- Too low or too high memory usage.
- Inappropriate disk activity.
- Unexpected frequency of file changes as measured for example by checksums or file size.
- SIGABRT signal from abnormally terminated UNIX processes.
- Presence of non-self.

Of course, it would also be possible to use 'positive' signals, as discussed in the previous section, such as the absence of some normal 'health' signals.

Consequently, those antibodies or detectors that match (first signal) those antigens within a radius, defined by a measure such as the above (second signal), will proliferate. Having thereby identified the dangerous components, further confirmation could then be sought by sending it to a special part of the system simulating another attack. This would have the further advantage of not having to send all detectors to confirm danger. In conclusion, using these ideas from the Danger Theory has provided a better grounding of danger labels in comparison to self/non-self, whilst at the same time relying less on human competence.

Conclusion

To conclude, the Danger Theory is not about the way AIS represent data. Instead, it provides ideas about which data the AIS should represent and deal with. They should focus on dangerous, i.e., interesting data.

It could be argued that the shift from non-self to danger is merely a symbolic label change that achieves nothing. We do not believe this to be the case, since danger is a grounded signal, and non-self is (typically) a set of feature vectors with no further information about their meaning.

The danger signal helps us to identify which subset of feature vectors is of interest. A suitably defined danger signal thus overcomes many of the limitations of self–non-self selection. It restricts the domain of non-self to a manageable size, removes the need to screen against all self, and deals adaptively with scenarios where self (or non-self) changes over time.

The challenge is clearly to define a suitable danger signal, a choice that might prove as critical as the choice of fitness function for an evolutionary algorithm. In addition, the physical distance in the biological system should be translated into a suitable proxy measure for similarity or causality in an AIS. We have made some suggestions in this paper about how to tackle these challenges in a variety of domains, but the process is not likely to be trivial. Nevertheless, if these challenges are met, then future AIS applications might derive considerable benefit, and new insights, from the Danger Theory.

Acknowledgments We would like to thank the two anonymous reviewers, whose comments greatly improved this paper.

References

1. Matzinger P (1998) An Innate sense of danger, Seminars in Immunology, pp 399–415
2. Matzinger P (2001) The danger model in its historical context. Scand J Immunol 54:4–9
3. Matzinger P (1994) Tolerance danger and the extended family. Annu Rev Immunol 12:991–1045
4. Burgess M (1998) Computer Immunology. Proc LISA XII:283–297
5. Williamson M (2002) Biologically inspired approaches to computer security, HP labs technical reports HPL-2002, pp 131
6. Somayaji A, Hofmeyr S, Forrest S (1998) Principles of a computer immune system. In: Proceedings New Security Paradigms Workshop, Charlottesville, pp 75–82
7. Aickelin U, Cayzer S (2002) The danger theory and its application to artificial immune systems. In: Proceedings of the 1st internet conference on artificial immune systems. ICARIS, Springer, pp 141–148
8. Aickelin U, Bentley P, Cayzer S, Kim J, McLeod J (2003) Danger theory: the link between AIS and IDS? In: Proceedings ICARIS-2003, 2nd international conference on artificial immune systems. ICARIS, Springer, pp 147–155
9. Goldsby R, Kindt T, Osborne B (2000) Kuby immunology, 4th edn. W H Freeman, New York
10. Bretscher P, Cohn M (1970) A theory of self–nonself discrimination. Science 169:1042–1049
11. Matzinger P (1994) Tolerance, danger and the extended family. Annu Rev Immunol 12:991–1045
12. Sundaram A (1996) An introduction to intrusion detection, crossroads: the ACM student magazine 2(4)
13. Forrest S, Perelson A, Allen L, Cherukuri R (1994) Self non-self discrimination in a computer. In: Proceedings of the 1994 IEEE symposium on research in security and privacy, pp 202–212
14. Forrest S, Hofmeyr S, Somayaji A, Longstaff T (1996) A sense of self for unix processes. In: Proceedings of the 1996 IEEE symposium on research in security and privacy, pp 120–128
15. Hofmeyr S, Forrest S (2000) Architecture for an artificial immune system. Evolutionary Computation 8(4):443–473
16. Somayaji A, Forrest S (2000) Automated response using system-call delays. In: Proceedings of the ninth USENIX security symposium, pp 185–197