Active Trust Management for Adaptive Survivable Systems (ATM for ASS's)

Howard Shrobe, Jon Doyle & Peter Szolovits

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: Computer systems are under attack as they have never been before. Every day computer systems around the world are compromised: data is stolen, security is breached, services are denied. These attacks are enabled by flaws in our current operating system, application code and middleware. It would be nice to replace all this with new software that is based on provably impenetrable trusted computing base, but of course this isn't practical. We instead need a way to make systems behave in a more survivable fashion and to complete their critical tasks even when compromises have happened.

Motivation: There are two ways to look at this problem of Information Survivability. One can either, as is traditional, try to build impenetrable walls around computer systems, or one can try to build highly adaptive computer systems. Such adaptive systems will find alternative ways of providing useful services that avoid using compromised resources as much as is possible. We believe that this unexplored approach is vital.

Previous Work: Most work that has looked at computer security has focused on either preventing intrusions or on rapidly detecting them so as to prevent damage. To date, there has been very little work on using adaptivity as the organizing principle behind survivability. We draw base our work on diagnostic techniques [1, 2] and techniques for software description [3].

Approach: Our project aims to build Adaptive Survivable Systems that are capable of performing their intended function even when underlying computational resources have been successfully compromised. In particular, we wish to build systems that model the trustworthiness of computational resources and that make rational choices about how best to achieve their goals in light of the risks and benefits involved in using alternative computational resources.

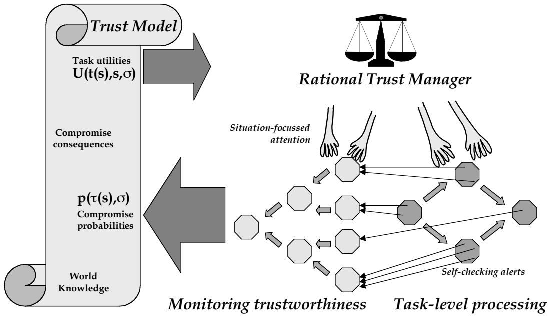

Figure 1: Automatic Trust Management for Survivability

Our project focuses on four major topics:

- Trust Models: An Adaptive Survivable System must know what resources are trustable and for what purposes they may be trusted. This in turn depends on what components have been compromised and on the form of the compromise. Finally, this depends on what what vulnerabilities exists, what attacks have been exploited these vulnerabilities, and with what intent they have been conducted. Our trust model therefore has three levels, each with its own ontology and inference techniques. The Trustability level centers on properties of significance to applications (e.g. privacy, quality of service). The compromise level focuses on computational components that provide these properties and on the ways in which they may be compromised. The attack level focuses on the types of attacks and on how they enable compromise of critical resources.

- Perpetual Analytic Monitoring: The trust model is constructed and kept current by constant monitoring of information streams arising from sources such as intrusion detection systems and the self-monitoring of application systems. We collate and analyze these reports, looking for temporal trends that are indicative of coordinated attacks or of particular compromises. This part of our effort will be based on our MAITA monitoring system.

- Self-Adaptive Survivable Systems: Trust models influence the way a Self-Adaptive system attempts to perform its computation. Self-Adaptive systems are structured so that each sub-task has many methods available for achieving its goal. Each of these methods requires specific types of resources and each of these resources is assessed for its trustworthiness; each method also promises a certain quality of answer. A self-adaptive system makes the rational choice of using that method which is most likely to achieve maximum net benefit. Self-adaptive systems also inform the trust model. The goals and invariants of each computation are explicitly represented and checked as the computation proceeds. The failure of a computation to behave as expected provides evidence that the resources used by that computation have been compromised. This evidence is reported to the monitoring system and is used to help assess the degree of compromise and the trustworthiness of the resources.

- Rational, Trust Driven Resource Allocation. Trend detection, self-monitoring and trust assessment all consume resources which might otherwise be used by applications to perform their critical services. Dedicating too many resources to house-keeping functions would prevent the applications from rendering their functions (i.e. a self-inflicted denial of service); dedicating too few resources to the house-keeping functions necessary for an accurate trust-model can lead to the use of compromised resources in tasks for which they are not trustworthy. Similarly, application systems themselves constantly make decisions about how to achieve their goals and which resources to use. Each of these decisions can be viewed as a rational decision making problem, that is assessing how best to achieve maximum expected net benefit, given the trustability of the resources, the political situation and the likelihood of coordinated, malicious intention.

Difficulty: Each of the major components of our approach involves significant research into uncharted waters. Our approach involves making rational decisions in real-time; but rational decision making is inherently unbounded. Thus we need to find more tractable, rule-based approaches to this decision making that adequately approximate the outcomes of decision theory.

Impact: This project is attempting to create an entirely new approach to the management of survivability in critical computational systems. We believe that current approaches are inherently limited and that approaches like ours are critically necessary.

Future Work: We have already constructed early prototypes of our diagnostic and monitoring components. We have also developed ontologies of attacks, vulnerabilities and compromises that form the core of the trust model. We plan to enhance our MAITA monitoring system to understand the information provided by a variety of intrusion detection systems and by self-monitoring applications. We are constructing a library of ”trend-templates” that describe the temporal pattern of behavior that characterize successful attacks and compromises. We are developing techniques for instrumenting an application system so that it checks its own progress towards. Finally, we are developing initial models for rational resource management that take into account the information in the trust model.

Research Support: This work is funded by a DARPA contract N66001-00-C-8078, administered by SPAWAR.

References:

[1] R. Davis, H. Shrobe, W. Hamscher, K. Wieckert, M. Shirley, and S. Polit. Diagnosis based on descriptions of structure and function. In *Proceedings of the AAAI National Conference on Artificial Intelligence*, pages 137–142, 1992.

[2] Johan de Kleer and Brian C. Williams. Diagnosing multiple faults. *AI*, 32:97–130, 1987.

[3] C. Rich, H. E. Shrobe, and R. C. Waters. An overview of the programmer's apprentice. In *IJCAI79*, 1979.