DOI:10.1145/2556938

William Young and Nancy G. Leveson

Inside Risks

An Integrated Approach to Safety and Security Based on Systems Theory

Applying a more powerful new safety methodology to security risks.

DESPITE AN ENORMOUS amount of effort and resources applied to security in recent years, significant progress seems to be lacking. Similarly, changes in engineering are making traditional safety analysis techniques increasingly less effective. Most of these techniques were created over 50 years ago when systems were primarily composed of electromechanical components and were orders of magnitude less complex than today's software-intensive systems. New, more powerful safety analysis techniques, based on systems theory, are being developed and successfully used on a large variety of systems today, including aircraft, spacecraft, nuclear power plants, automobiles, medical devices, and so forth.[2] Systems theory can, in the same way, provide a powerful foundation for security. An additional benefit is the potential for creating an integrated approach to both security and safety.

The Relationship Between Safety and Security

Practitioners have traditionally treated safety and security as different system properties. Both communities generally work in isolation using their respective vocabulary and frameworks. Safety experts see their role as preventing losses due to *unintentional* actions by *benevolent* actors. Security experts see their role as preventing losses due to *intentional* actions by *malevolent* actors. The key difference is the intent of the actor that produced the loss event. It may never be possible to determine this intent—but if the majority of our energy and analysis is refocused on building better loss prevention strategies (regardless of actor intent), then it may not matter. We are not suggesting that intent need not be considered, only that the problem can be reframed as a general loss prevention problem that focuses on the aspects of the problem (such as the system design) that we have control over rather than immediately jumping to

CREDIT TK

the parts about which we have little information, such as identifying all the potential external threats.

Note the common goal of mission assurance here, that is, the ability to complete a mission while enforcing constraints on how the mission can be achieved. In a nuclear power plant, for example, the goal is to produce power while preventing the release of radioactivity. The causes for not producing the power or for releasing radioactivity may be due to accidental or malicious reasons, but the high-level goal of preventing these events is the same.

By taking a common top-down, system engineering approach to security and safety, several benefits accrue. One is that the overall role of the entire socio-technical system as a whole in achieving security and safety can be considered, not just low-level hardware or operator behavior. Others include more efficient use of resources and the potential for resolving conflicts between safety and security early in the development process.

Applying systems theory and systems engineering to security requires initially focusing security on high-level strategy rather than immediately jumping to the tactics problem. Certainly adversary action is a critical consideration in addressing security and preventing intentional losses. Yet, focusing on adversaries or threats too early in the process, absent the benefit of context, limits the overall strategic-level utility of the security assessment. Stated another way, the goal of security is not to guard the physical network and prevent intrusions, which is threat focused. The goal is to ensure the critical functions and ultimately the services that the network and systems provide are maintained in the face of disruptions. By changing to a strategic viewpoint rather than starting with tactics, security analysts and defenders can proactively shape the situation by identifying and controlling system vulnerabilities rather than defending from a position of disadvantage by being forced to react to continually changing threats and other environmental disruptions.

Tactics are prudent means to accomplish a specific action (such as guarding networks and other information assets).

Coming Next Month

Strategy vs. Tactics in Security

The security field tends to draw heavily on language, metaphors, and models from military operations. As a result, much of cybersecurity is typically framed as a battle between intelligent, adaptive adversaries and defenders. Security focuses on how defenders can close holes in their networks that might otherwise allow adversaries to gain access and create disruptions. Defenders apply best practices (tactics) in order to protect the network and other information assets.

There is an important distinction between tactics and strategy. Strategy can be considered as the art of gaining and maintaining continuing advantage. On the other hand, tactics are prudent means to accomplish a specific action (such as guarding networks and other information assets). Tactics is focused on physical threats, while strategy is focused on abstract outcomes.

In tactics models, losses are conceptualized as specific events *caused* by threats. For example, a security incident consisting of a data breach with an accompanying loss of customer Personally Identifiable Information (PII) is viewed as a single occurrence, where an adversary successfully precipitates a chain of events leading to a loss. The chain of events typically translates into attackers successfully negotiating several layers of defenses such as firewalls and encryption. In almost all such cases, security analysts will identify some proximate cause that should have served as the last barrier or line of defense. If only the barrier would have been in place, then the attack would have failed. Although threats exploiting vulnerabilities produce the loss event, tactics models treat the threat as the *cause* of the loss.

Preventing losses, then, is heavily dependent on the degree to which security analysts can correctly identify potential attackers—their motives, capabilities, and targeting. Once equipped with this knowledge, security experts can analyze their systems to determine the most likely route (or causal chain) attackers may take to achieve their goal. Resources can then be allocated to erect a "defense in depth" to prevent losses.

Threat prioritization is also challenging given the sheer volume of threats. If the defense is optimized against the wrong threat, then the barriers may be ineffective. Perhaps an unstated assumption is that defense against the more sophisticated threats can handle so-called lesser-included cases, but this is not necessarily the case. Simple requirements errors or operational procedures may allow even unsophisticated attacks from previously ignored or lower-level adversaries to succeed.

In contrast to a tactics-based, bottom-up approach, a top-down, *strategic* approach starts with identifying the system losses that are unacceptable and against which the system must be protected. The result is a

small and more manageable set of potential losses stated at a high-level of abstraction. These losses likely extend beyond the physical and logical system entities into the higher-level services provided by these entities.

Rather than starting with the tactics questions of *how* best to guard the network against threats, a strategic approach begins with questions about *what* essential services and functions must be secured against disruptions and *what* represents an unacceptable loss. The "whats" will be used later to reason more thoroughly about only the "hows" that can lead to specific undesirable outcomes. The analysis moves from general to specific, from abstract to concrete. (Robinson and Levitt[5] similarly considered abstraction layers with respect to being able to prove emergent system properties hierarchically.)

One of the most powerful ways human minds deal with complexity is by using hierarchical abstraction and refinement. By starting at a high level of abstraction with a small list and then refining that list with a more detailed list at each step (working top down), one can be more confident about completeness because each of the longer lists of causes (refined hazards or causes) can be traced to one or more of the small starting list (and vice versa).

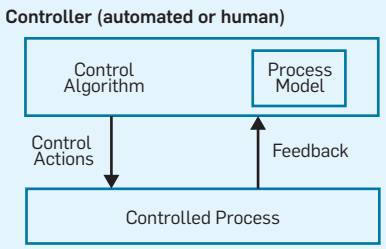
With traceability, it is also easier for human reviewers to find any incompleteness. We say "more confident" because such a list can never be proven to be complete—there is no formal (mathematical) model of the entire system and how it will operate. Human participation in the analysis and human review of the results will always be required and, therefore, incompleteness will always be possible. But structuring the process in a way that optimizes human processing and review will reduce any potential incompleteness.

Focusing first on strategy rather than tactics can be achieved by adopting a new systems-theoretic causality model recently developed to provide a more powerful approach to engineering for safety.

A New Systems-Theoretic Approach to Security and Safety

The limitations of traditional engineering methods and the need to field increasingly complex systems during and immediately following World War II led to the development of modern systems theory in the 1940s and 1950s.[1] Systems theory provides the philosophical and intellectual foundation for systems engineering and for a new, more inclusive model of accident causality called STAMP (System-Theoretic Accident Model and Processes).[2]

Figure 1. A basic control loop.



Traditional causality models used in safety attribute accidents to an initial component failure or human error that cascades through a set of other components. One way to envision this model is as a set of dominoes. At one end is the initial domino, which is representative of a single human error or component failure. This initial error is labeled as the root cause. The failure propagates through the system, leading to the failure of other components until the last domino falls and the loss occurs. In this model, the first domino causes the last domino to fall (the actual loss event). Moreover, if any of the intervening dominoes are removed, the chain is broken.

This model is effective for systems with limited complexity, for example, linear interactions and simple cause-and-effect linkages like dominos (or holes in Swiss cheese, another common analogy).

Today's increasingly complex, software-intensive systems, however, are exhibiting new causes of losses, such as accidents caused by unsafe interactions among components (none of which may have failed), system requirements and design errors, and indirect interactions and systemic factors leading to unidentified common-cause failures of barriers and protection devices. Linear causality models and the tools built upon them, like fault trees, simply lack the power to include these new causes of losses.

STAMP is a new systems-theoretic model of causality related to emergent system properties. It was originally created to act as a foundation for more powerful approaches to safety. Security, however, is also an emergent system property, and STAMP and its associated analysis tools are equally applicable to security. STAMP envisions losses as resulting from interactions among humans, physical system components, and the environment that lead to the violation of safety constraints. The focus shifts from "preventing failures" to "enforcing safety constraints on system behavior." While enforcing safety constraints may require handling component failures, other inadvertent and advertent causes must also be controlled.

Constraints on system behavior are enforced by controls in a hierarchical control structure, where each level of the structure enforces the required constraints on the behavior of the components at the next lower level. Control loops operate between each level of this control structure, with control actions shown on the downward arrows and feedback on the upward arrows. Figure 1 shows the general form of such control loops. In both safety and security, the goal is to prevent (constrain) control actions that can lead to losses under worst-case environmental conditions.[3]

In systems and control theory, every controller must contain a model of the process it is controlling. This model is used to determine what control actions are necessary. Many accidents related to software or human operators are not the result of software or human "failure" (whatever that might mean), but instead stem from inconsistencies between the controller's models of the controlled process (usually called a mental model for human controllers) and the actual process state. For example, friendly fire accidents are usually the result of thinking a friendly aircraft is an enemy and executing unsafe control actions. Whether the inconsistency results from an inadvertent reason (accidental loss of feedback, for example) or tricking the controller into thinking that the friendly aircraft is an enemy (purposeful creation of incorrect feedback), the result remains the same—an unsafe or unwanted control action.

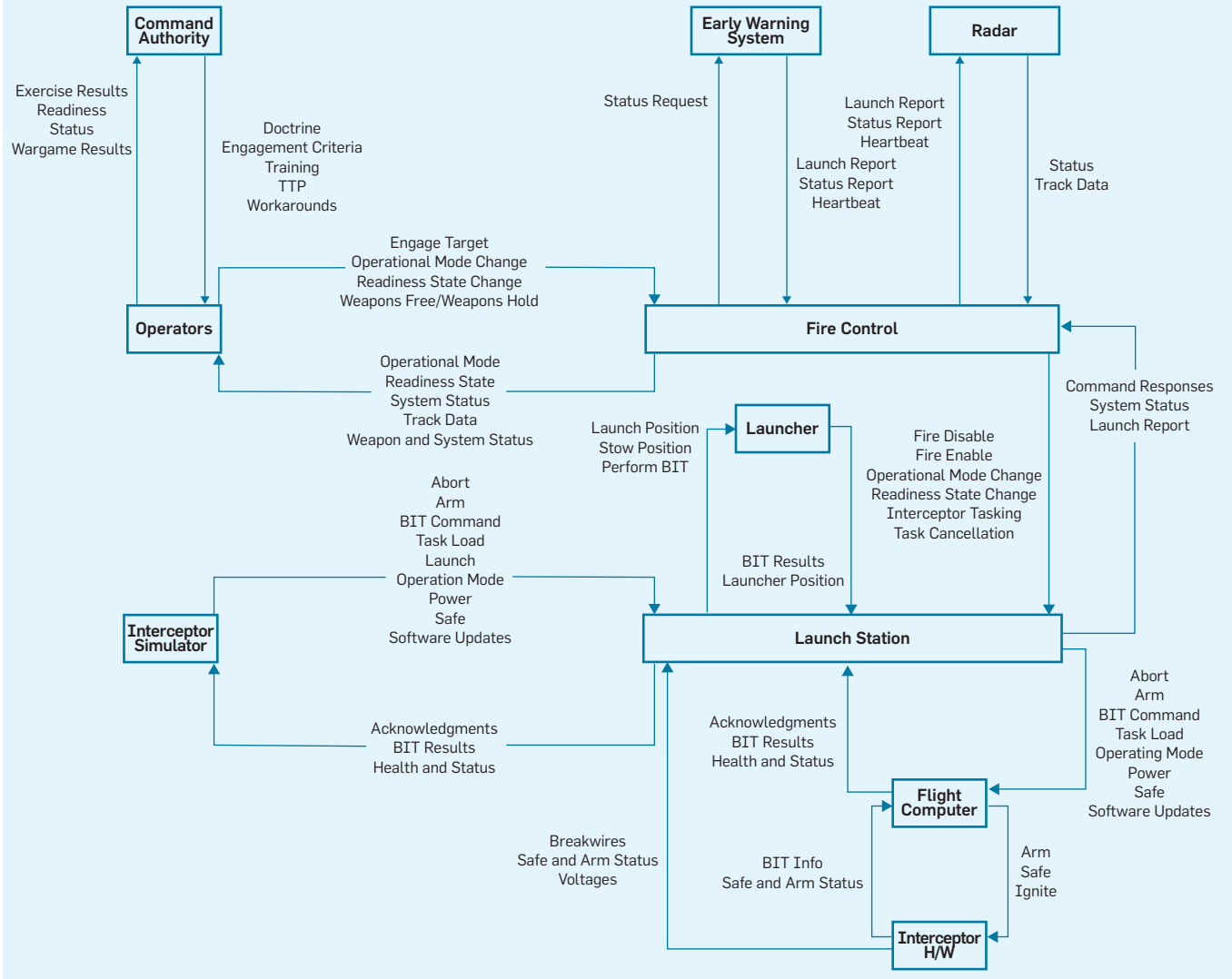
Stuxnet provides another example. The automated system (controller) thought the centrifuges (controlled process) were spinning at a slower speed than they actually were, and issued an *Increase Speed* command when the centrifuges were already spinning at maximum speed, which led to equipment damage. (A loss that officials probably wanted to prevent.)

New and more powerful techniques for safety analysis and design have been created on this theoretical foundation. STPA (System-Theoretic Process Analysis), for example, is a new hazard analysis technique based on the STAMP model of causality. The analysis is performed on the system functional control structure. Figure 2 depicts an illustrative functional control structure for a ballistic-missile defense system.[2,4] In this example, there are several safety and security critical control commands, such as *fire enable* and *launch interceptor*.

One key point worth emphasizing is the fact that the function control model contains physical aspects, social aspects, logical and information aspects, operations and management aspects. Performing the hazard (safety) or vulnerability (security) analysis on such a model allows a broad perspective on potential causes for a loss. Most hazard and vulnerability analysis techniques use physical system models rather than functional system models, and thus concentrate on physical component failures rather than dysfunctional (unsafe or insecure) system behavior and broader social and organizational factors.

Once the control structure is created, the first step in the STPA analysis is to identify potentially unsafe control actions, which in general include (1) providing a control action that leads to a hazard (for example, a missile is launched at a friendly aircraft), (2) not providing a control action that is needed to prevent a hazard (for example, a missile is not launched to down an enemy aircraft), (3) providing a control action too early or too late or out of sequence (for example, a missile is launched but too early or too late to be effective in preventing a loss), or (4) continuing a control action too long or stopping it too soon. Losses can also result from a safe (required) control action that is not executed properly (for example, the launch missile instruction is not executed correctly). After

Figure 2. Functional control structure for a ballistic missile defense system.



the unsafe control actions have been identified, the second step involves examining the system control loops (using a structured and guided process) to identify scenarios that can lead to the identified unsafe control actions.

STPA-Sec is an extension to STPA to include security analysis. The initial steps in the analysis are identical to those for safety: identifying the losses to be considered, identifying system hazards or security vulnerabilities, drawing the system functional control structure, and identifying unsafe, or in this case, insecure, control actions. The only difference is the addition of intentional actions in the generation of the causal scenarios, the last step in the process.

STPA is currently being used on safety problems in a wide variety of industries. Careful evaluations and comparisons with traditional hazard analysis techniques have found that STPA finds the loss scenarios found by the traditional approaches (such as Fault Tree Analysis and Failure Modes and Effects Analysis) as well as many more that do not involve component failures. Surprisingly, while STPA is more powerful, it also appears to require fewer resources, including time.

STPA-Sec is only now being applied to cybersecurity problems, but is showing promise in these case studies. A formal evaluation and comparison with real red teams using traditional security analysis techniques such as attack trees will be completed by spring 2014.

Another benefit of using a tool based on a system-theoretic model is that it can be applied earlier in the design process and in situations where specific component data is unavailable. Analysis can begin as soon as the basic high-level goals (mission) of the system is identified and design decisions evaluated for their impact on safety and security before expensive rework is necessary. As the detailed design decisions are made and the design refined, the STPA/STPA-Sec analysis is refined in parallel.

Conclusion

By using a causality model based on systems theory, an integrated and more powerful approach to safety and security is possible. Hazards lead to safety incidents in the same way that vulnerabilities lead to security incidents. We argued in this column that the key question facing security analysts should be how to control vulnerabilities, not how to avoid threats. Rather than initially trying to identify all the threats and then move up to the vulnerabilities they might exploit to produce a loss, a top-down systems engineering approach starts with system vulnerabilities, which are likely far fewer than threats and, if controlled, can prevent losses due to numerous types of threats and disruptions. This top-down approach also elevates the security problem from guarding the network to the higher-level problem of assuring the overall function of the enterprise.

The key question facing security analysts should be how to control vulnerabilities, not how to avoid threats.

Use of a systems-theoretic approach to security, however, requires a reframing of the usual security problem. Just as STAMP reframes the safety problem as a control rather than a failure problem, applying STAMP to security involves reframing the security problem into one of strategy rather than tactics. In practice, this reframing involves shifting the majority of security analysis away from guarding against attacks (tactics) and more toward design of the broader socio-technical system (strategy). Put another way, rather than focusing the majority of the security efforts on threats from adversary action, which we have limited control over, security efforts should be focused on the larger, more inclusive goal of controlling system vulnerabilities.

Controlling vulnerabilities allows security analysts to prevent not only disruptions from known threats, but also disruptions introduced by unknown threats, such as insiders. In other words, the source of the disruption does not matter. What matters is identifying and controlling the vulnerabilities. This approach limits the intelligence burden required to perform the initial system security analysis. The analysis will eventually address threats, but does so much later in the process after generating a deeper systemic understanding of the context under which the threats may operate and the disruptions that actually lead to critical loss events.

Because contemporary security and safety both attempt to prevent losses in complex software-controlled systems, we believe applying the same system-theoretic causality model may benefit security the same way it is benefitting safety. Research is currently under way to test this notion. The key underlying idea is that from a strategy perspective, the physical (or proximate) cause of a disruption does not really matter. What matters is the efficacy of the strategy in dealing with (controlling) the effects of that disruption on overall system function or assuring the mission. This is a significant paradigm shift for security experts (as it was for safety experts). While likely to force a reexamination of many of the accepted truths of security, we believe such a refocus will help address three of the major problems with contemporary approaches to security—quantity, threat variety, and threat prioritization—can all be addressed more effectively through this new approach than through existing approaches. The new approach does not discard traditional security thinking, but does suggest it is tactically focused and must be augmented by an effective strategy in order to succeed. C

References

1. Checkland, P. *Systems Thinking, Systems Practice.* John Wiley & Sons, New York, 1981.
2. Leveson, N.G. *Engineering a Safer World*, MIT Press, 2012.
3. Leveson, N. and Thomas, J. *An STPA Primer*; http://sunnyday.mit.edu/STPA-Primer-v0.pdf
4. Pereira, S.J., Lee, G. and Howard, J. A system-theoretic hazard analysis methodology for a non-advocate safety assessment of the ballistic missile defense system. In *Proceedings of the 2006 AIAA Missile Sciences Conference* (Monterey, CA, Nov. 2006).
5. Robinson, L. and Levitt, K.N. Proof techniques for hierarchically structured programs. *Commun. ACM 20*, 4 (Apr. 1977), 271–283.

William Young (wyoung@mit.edu) is a Ph.D. candidate in the Engineering Systems division at Massachusetts Institute of Technology, Cambridge, MA.

Nancy G. Leveson (leveson@mit.edu) is Professor of Aeronautics and Astronautics and also Professor of Engineering Systems at Massachusetts Institute of Technology, Cambridge, MA.