DOI:10.1145/2988445

Peter G. Neumann

Inside Risks

Risks of Automation: A Cautionary Total-System Perspective of Our Cyberfuture

Where automation is inevitable, let's do it right.

Many computer-related risks discussed in past Inside Risks columns are still present today. These risks (and new ones) are likely to intensify even further as systems provide extensive automated or semi-automated operation. Significantly greater total-system trustworthiness will be required, encompassing better hardware, system software, and applications that are able to tolerate human limitations and environmental factors. Risks will continue to result from inadequate reliability, security, and privacy, as well as gullibility and general inability of users to cope with complex technology. We repeatedly discover unexpected risks resulting from lashing subsystems together (for example, see Beurdouche[2]), because of unexpected system behavior. Many advances in research, system development, and user friendliness are urgently needed. Also, some middle ground is desirable between the optimists (who believe there are easy answers to some of the problems posed here) and the pessimists (who have serious doubts about increasing uses of automation and artificial intelligence—especially when used by people who are more or less technologically queasy).

In this column, I examine certain approaches that might be economically desirable, but that have serious potential risks. These include aviation safety and security; self-driving and semi-automated vehicles, and eventually automated highways; the so-called Internet of Things; and cloud computing and cloud storage.

Total-system trustworthiness must recognize requirements for human safety, security, reliability, robustness, and resilience despite adversities such as human error, attacks, and malware. However, we also need proactive system architectures that inherently minimize the extent to which various components have to be trusted, and other requirements such as extensive monitoring, auditability, interoperability, compatibility, and predictable composability of components to enable facile multi-vendor systems. For example, voice and speech recognition and understand-

Lorem ipsum dolor sit amet consect adipiscing nunc enim mauris sed massa

CREDIT TK

ing, automatic translation, intelligent dialogues, and automated responses have some potentials to compromise trustworthiness. Also, we must depend upon systems and networks that are intrinsically untrustworthy in various respects—and sometimes made even less so by human frailty, insider misuse, and potential governmental desires for exceptional accesses that bypass already marginal security (for example, see Abelson et al.[1]). As a result, we need people-tolerant systems as well. Above all, we will need scalability of the implementations with respect to all of the requirements mentioned here (whether or not individual local control is also desired), plus the inevitable desire for remote upgrades to quickly remediate system vulnerabilities and to enable new applications. All of this is very daunting in light of the reality that we are trying to evolve incrementally from today's flaky platforms. Thus, we might wonder whether some of these desiderata are actually pipedreams that cannot be implemented, maintained, and used with sufficient assurance that the remaining risks will be acceptable. No system is ever going to be perfect—especially ones that require considerable autonomy in operation. However, the question of what is good enough always remains; it cannot be answered generally, largely because there are different answers depending on the specific applications.

Aviation Safety and Security

We are already pushing the edges with regard to aviation safety and security in the large. Developing avionic system hardware and software that cannot be subverted accidentally or intentionally is demonstrably nontrivial and expensive, but only a small part of the overall problem. This was originally conceived as the *Free-Flight* program, putting much greater smarts in cockpit control systems—so that air-traffic controllers on the ground might become less critical in real time. For example, collision-avoidance systems are now well established and generally reliable. Free-Flight has now morphed more generally into the total-system *NextGen* program, which will integrate ground- and air-based controls. However, the notion of having safe distributed heavily automated control

No system is ever going to be perfect—especially ones that require considerable autonomy in operation.

among nearby aircraft in the broader context of airport and long-range enroute scheduling, with real-time total traffic control (especially in times of inclement weather delays) could introduce many potential risks. In that air-traffic controllers and pilots today may be sorely pressed in times of heavy congestion and erratic weather conditions, providing them with more intelligent computer-aided relief should be beneficial—if it can be assuredly provided. For example, the new DO-178C certification tool suite has evolved significantly, and is considerably more advanced than its predecessors. It offers significant hopes that we can further increase flight safety and security.

Aviation safety and security are of course a worldwide concern, not just a domestic one, especially with many different countries and languages—and problems requiring emergency remediation. Enormous progress has been made along these lines, although there are still corner cases that may defy adequate control and require pilot attention (and possible intervention). However, putting most of the controls in the hands of integrated automation must encompass hardware, software, communications, pilots who might or might not be able to override computer controls in emergencies, ground controllers with excellent training and experience, and defenses against would-be intruders. Infotainment systems have tended to coexist on the same local network with the aircraft controls, perhaps without adequate separation. The total-system approach must therefore develop stronger network security to ensure that the flight-control systems are strongly isolated from the infotainment and systems.

Other problems within the total-system perspective include airport safety and security, passenger screening, timely preventive aircraft maintenance, and thorough pilot training that anticipates unexpected events. We tend to put our eggs in a few defense mechanisms (including those that were not previously present to thwart past compromises); however, that is not a viable strategy when there are too many vulnerabilities.

It is also necessary to consider the presence of remotely controlled drones sharing the air space, and all of the risks to human safety and privacy, in flight and on the ground. Drones (mostly semi-autonomously or manually controlled at present, although they could be fully autonomous in the future) will require better security to prevent subversion akin to that demonstrated in modern automobiles—particularly, drones carrying lethal weapons.

Automotive Safety in Automated Vehicles

Total-system safety and security concerns include the demonstrated ability to compromise the controls of conventional vehicles—for example, through the wireless maintenance port or otherwise gaining access to the internal local network. Those problems must be addressed in vehicles with self-driving or highly automated features. Note that a distinction is made here between self-driving cars (for example, Google, albeit with a surrogate driver during the current test and evaluation phases, but with the intention of becoming fully autonomous) and computer-augmented driver assistance (for example, Tesla) that goes way beyond more familiar features such as cruise control, airbags, anti-lock braking, parallel parking, rear-vision video, and other recent enhancements for safety and convenience, but that falls somewhat short of fully autonomous control with no ability for manual intervention.

Those of us who live in the California Bay Area frequently encounter self-driving Google cars. The accident rates thus far are very low, in part because the vehicles are programmed to aggressively observe traffic signs and environmentally changing road conditions—usually with the surrogate driver ready to override. (There are

cases of Google vehicles being hit from behind by human drivers—primarily because of Google's conservative programming; it is thought that the cars running into them may be following too closely, with drivers who are not cognizant of the conservative nature of the Google car.) The desires for dramatically reducing accident rates through vehicle automation seem realistic, although there are always likely to be unanticipated corner cases. Incidentally, Google has monitored some of the surrogate drivers, and discovered they tended not to be paying strict enough attention—perhaps because the vehicles performed so well! In any case, the record of self-driving Google vehicles seems vastly better than that of old-fashioned human-driven ones. Recognizing that the evolving automation is still a work in progress, there is considerable hope.

Unfortunately, the "driver" of a Tesla S died on May 7, 2016, in a crash in Florida while his car was in the automated-assistance mode.[a] This is reportedly the first known fatal accident involving a vehicle under automated control. Joshua Brown (a Navy veteran who had founded his own technology consulting firm) was in the driver's seat with no hands on the steering wheel, and was an outspoken advocate of the safety of the automated controls. (Recent reports suggest that he was watching a Harry Potter movie.) The cited article states that "Neither the Autopilot nor the driver noticed the white side of a tractor-trailer [which made a left turn in front of the Tesla] against a brightly lit sky, so the brake was not applied." The crash seems to cast doubts on whether autonomous vehicles in general can consistently detect all potential life-threatening situations. However, after a reported million miles of driving, a single fatality may not be particularly significant. This is far better than human driving. Although the details raise concerns, even seemingly perfect automation would still lead to accidents, injuries, and deaths; even with automation, nothing is actually perfect.

Karl Brauer (a Kelley Blue Book analyst) was quoted: "This is a bit of a wake-up call. People were maybe too aggressive in taking the position that we're almost there, this technology is going to be in the market very soon, maybe need to reassess that." However, Elon Musk has praised the Tesla Model S as "probably better than a person right now." Also, a Tesla statement on June 30 noted that driving a Model S with this technology enabled (as a beta-tester!) "requires explicit acknowledgment that the system is new technology."

Recognizing that the evolving automation is still a work in progress, there is considerable hope.

An immediate reaction to the Tesla "Autopilot" is that it should not be called an *autopilot*, because it explicitly demands constant attention from the person in the driver's seat. This misnomer has been raised repeatedly—especially in the aftermath of the recent accidents.

The Tesla involved in Brown's death did not have LIDAR (Light Detection and Ranging) pulsed lasers, and was relying on the Mobileye camera and forward-facing radar.[b] It is clear that many improvements can be added (such as LIDAR)—not just to the vehicle controls, but also by automating the sensors and signals in roadways and particularly in dangerous intersections themselves, dynamically establishing different speed limits under bad weather conditions, and much more.

On July 6, 2016, reports appeared that a Tesla X on "Autopilot" lost control on the Pennsylvania Turnpike, bounced off concrete guard rails, and flipped over; the passenger in the driver's seat was reportedly not paying enough attention, and was injured.[c]

John Quain[7] notes there is significant evidence that a driver behind the wheel may not be ready to take over from the autopilot quickly enough to avert a disaster: "Experiments conducted last year by Virginia Tech researchers and supported by the national safety administration found that it took drivers of Level 3 cars [in which the driver can fully cede control of all safety-critical functions in certain conditions] an average of 17 seconds (!!!) to respond to takeover situations. In that period, a vehicle going 65 mph would have traveled 1,621 feet—more than five football fields."[d]

Generalizing this situation, a huge question seems to arise regarding liability—where litigation tends to look for deep pockets. But there are many issues here. Perhaps when you buy an automated vehicle, the contract might stipulate that the car is *experimental* and that the maker disclaims liability, explicitly waiving responsibility. (This is somewhat akin to the providers of the most common operating systems declaring that these systems should not be used for critical applications—although that caveat seems to be widely ignored.) In that case, the maker's lawyers might successfully claim that a driver was negligent by having too much faith in the software/hardware system. The legal issues are further complicated if highway patrols insist on backdoors to be able to redirect or stop vehicles for inspection or arrest, which itself might cause an accident or a violent action. And what happens when two or more totally driverless autonomous vehicles actually collide? Or when a remotely controllable vehicle is coopted for evil purposes? There are vastly too many risks to enumerate here, and much more research, development, and evaluation are needed.

In particular, consider two highly relevant papers by Don Norman[3,6] well worth reading. Don contributed some pithy quotes for my ACM *Ubiquity* July 2016 article[4] on this subject. He believes that partial automation is a disaster waiting to happen, and that total automation is essential. "To think otherwise is to ignore decades of solid research from the psychology and human factors fields (and the National Academy's Human Systems Integration board). And there is no way

a See http://bit.ly/2aRzPqX

b See http://bit.ly/297eo4D

c See http://bit.ly/2aYNzBD

d Vlasic, B. and Boudette, N. *The New York Times* (July 1, 2016), with follow-up posts by Bill Vlasic the following day and week; http://nyti.ms/2b2QC91

to overcome it. The better the [partial] automation, the more dangerous it becomes. It has to be full automation, not this silly Level 3."

However, introducing automation into activities already regulated by standards that were not formulated with automation and security in mind can introduce risks. Also, lack of infrastructural investment and demands for incremental change with backward compatibility may be impediments to progress toward safety and security.

While writing this column, I learned of the Automotive Information Sharing and Analysis Center (Auto-ISAC, which has assembled a set of best practices) and The Billington Global Automotive Cybersecurity Summit (which had its inaugural meeting on July 22, 2016). These efforts seem to echo my concern that safety and security must be considered together throughout the automotive industry. Indeed, they claim to do so without seeking to make security a competitive advantage for individual companies, to learn what they can from other sectors, and to make fully autonomous cars available on an ordinary retail basis within the next 10 years.[e]

Automated Highways

The concept of every vehicle on a highway being automated (without fear of accidents or frustrations from congestion) still may seem somewhat remote. It will ultimately rely on highly collaborative coordination among neighboring vehicles in addition to the automation and semi-automated assists noted in the preceding section, and trustworthy communications with neighboring vehicle controllers and road hazards. In addition, some sort of total-system traffic monitoring is going to be essential, especially in detecting and responding to accidents, extreme weather conditions, vehicles running out of fuel or battery, flat tires, and more. Another concern is of course introducing older vehicles (with minimal autonomy and real-time monitoring) into the mix, or perhaps living with a simpler solution—barring such legacy vehicles from the automated highway and forcing them onto back roads. The two-dimensional control problems may be slightly less challenging than the three-dimensional aircraft flight control problems, but nonetheless important, particularly in potential emergencies. However, the separations among moving objects, the human vs. automated reaction times, and the ensuring risks differ widely among in-flight aircraft and ground vehicles. In some ways, the automation problem in aviation is simpler than that with automobiles. Pattern recognition for a variety of objects in confusing backgrounds is critical in automobiles; in airplanes, one simply has to detect the presence of an object—because the exact identity does not much matter (except in combat). Moreover, responses in driving may be needed within fractions of a second, whereas the time required in aviation is typically measured in minutes—or even hours for long-range anti-congestion planning.

The total-system concept applies acutely to automated highways, as many problems must be integrated. The entire environment may be laced with sensors and instrumentation that can interact with individual subsystems—signaling them appropriately in real time. This will create many complex interconnected system problems requiring scalable solutions and that avoid excessive energy consumption.

As a consequence, the legal, liability, privacy, and other issues noted here for automated vehicles are likely to be even more complicated when applied to distributed control of autonomous and semi-automated vehicles on automated highways, even if they are not co-mingling with conventional manual vehicles.

Any attempt to develop autonomous systems must have intensive monitoring to ensure that the systems are operating properly.

The Internet of Things

The Internet of Things (IoT) has the potential that almost everything imaginable might have some sort of online presence. Therefore, the IoT must be considered in the context of the preceding discussion—particularly with respect to those Things that are actually directly accessible on the Internet. Devices may be completely autonomous or operated totally under human control (but with remote monitoring), or again in between. Some will be remotely controllable, or otherwise accessible over the Internet. (This seems to be an open invitation for undesirable manipulation and invasive privacy violations.) However, more sensibly, many such Things are likely to be hidden behind a firewall, but still potentially accessible remotely (for example, via SSH). If the cheapest solutions are sought, there might be no firewall, and each Thing would require its own protective environment. Otherwise, the existing "Dark Net" on the Internet (generally unsearchable) will grow significantly to accommodate all of the Things that might hide behind supposedly secure firewalls. This could also result in the development of firewalls that are penetrable for government surveillance in certain countries, which could open up misuse by others as well. Given the vulnerabilities in today's firewalls, desktops, and mobile devices, significantly better security will be required, even in tiny systems in small and seemingly inconsequential Things, but especially in firewalls and internal routers. Indeed, perhaps those seemingly inconsequential ones will provide access to the others—because of the likelihood of unrestricted total access within the locally networked Things behind the firewall.

The privacy issues are somewhat murky. For example, a federal judge for the Eastern District of Virginia has ruled that the user of any computer that connects to the Internet should not have an expectation of privacy, because computer security is ineffectual at stopping hackers. The June 23, 2016, ruling came in one of the many cases resulting from the FBI's infiltration of PlayPen, a hidden service on the Tor network that acted as a hub for child exploitation, and the subsequent pros-

e "'Gene,' Tesla Model X rolls over after crashing into concrete divider, driver claims Autopilot was activated," (July 6, 2016); http://bit.ly/2bcr9KM and AFP news item, "Tesla crash: Model X flips while in autopilot mode, driver says"; http://bit.ly/2aMWmTO

ecution of hundreds of individuals. (The judge's ruling seems in conflict with other rulings, and could well be appealed.) To identify suspects, the FBI took control of PlayPen for two weeks and used a network investigative program that runs on visitors' computers to identify their Internet addresses.[f]

We might suspect today that the IoT is largely a corporate marketing opportunity where each company seeks to have a valid approach. However, it also appears that there is no *there* there—at least not yet, and that you might expect a lot of snake-oil salesmen.

Clouds

Cloud computing and cloud storage make enormous sense in many operational environments. To most users, these resources would seem to be autonomous, with human inputs and computer-generated outputs. However, they raise many issues relating to the trustworthiness of the clouds and networks, and who or what needs to be trusted. Examples of what might be particularly thorny here are encryption and key management, exceptional access for law enforcement, and maintenance and remediation when something goes fundamentally wrong (for example, outages or compromise). In the last of these concerns, where might you (or the cloud provider) find suitably experienced system administrators rapidly in cases of crises? Most of these issues may be completely out of the control of user communities.

Surveillance

The Keys Under Doormats report[1] makes the technical argument that dumbing down security to simplify the job of law enforcement is a very bad idea: for example, it would open up huge potential vulnerabilities for exploitation, and would undoubtedly drive domestic system providers and their domestic customers in many different nations to find other sources of secure systems. Several former high U.S. government officials have supported the conclusions of that report.

Any attempt to develop autonomous systems must have intensive monitoring to ensure that the systems are operating properly. As a consequence, the challenges of developing monitoring that is not only trustworthy, nonsubvertible, and privacy-aware, but also forensics-worthy will have to be addressed. The risks of dumbed-down security being compromised by other than the supposedly privileged surveillers (including privileged insiders) will add to the reality that automobiles and other devices could be remotely compromised. As a result, demands for surveillable autonomous systems that cannot be compromised by others seems to be an oxymoronic idea, or perhaps recursively difficult—as it would require much more secure systems in the first place!

We need computer-related systems with significantly greater trustworthiness.

Remediation

Some of these problems (except for "noncompromisible surveillance") can be addressed by having hardware that enforces fine-grained access controls along with hardware-ensured virtualization, and scalable compartmentalization of software that may be less trustworthy. For example, mobile devices and laptops should not allow applications to have unfettered access to contact lists and other apps without explicit permission. Hardware that helps enforce strict security properties would be very beneficial. Similarly, the Internet of Things will require seriously secure firewalls and local networks, with subsystems scaled in cost and complexity according to the criticality of the Things. Advances in formal methods can also play a role in increasing the assurance of trustworthiness of the hardware and software of such systems, including formally based testing and evaluation. See the CHERI system architecture[8] for an example of what might be possible with clean-slate hardware design, with operating system and compiler variants that know how to take advantage of the hardware.

Conclusion

Purveyors of modern computer-based systems wish to make some great leaps forward with automation and real-time automated assistance—in some cases bringing beta-test versions into use prematurely. We need computer-related systems with significantly greater trustworthiness than we have today, especially for use in critical systems. We also need much more stringent total-system requirements and overall system architectures, better development engineering, total-system testing and evaluation, and—perhaps above all—proactive awareness and understanding of the risks for would-be customers. If we are routinely going to have fully automated systems—or even partially automated systems that may require instantaneous human interventions in certain cases—we must have much more advanced system research and development, as well as education relating to potential risks and how to deal with them when they arise. The old adage "Let the buyer beware" (Caveat Emptor) must be extended to users as well. C

f See http://nyti.ms/2aHGExM

References

1. Abelson, H. et al. *Journal of Cybersecurity 1*, 1 (Nov. 2015), Oxford University Press; http://bit.ly/2bcj1dr
2. Beurdouche, B. et al. A messy state of the union: Taming the composite state machines of TLS. In *Proceedings of the 36th IEEE Symposium on Security and Privacy*, San Jose, CA, May 18–20, 2015; http://bit.ly/2bndXGz
3. Casner, S.M., Hutchinson, E.L., and Norman, D. The challenges of partially automated driving: Car automation promises to free our hands from the steering wheel, but might demand more from our minds. *Commun. ACM 59*, 5 (May 2016).
4. Neumann, P.G. Automated car woes—Whoa there! *ACM Ubiquity*, July 2016; http://bit.ly/2aYKDoT
5. Neumann, P.G. *Computer-Related Risks.* Addison-Wesley and ACM Press, 1995.
6. Norman, D.A. The human side of automation. *Road Vehicle Automation 2*, Springer, 2015.
7. Quain, J.B. The autonomous car vs. human nature, a driver behind the wheel may not be ready to take it. *The New York Times* (July 8, 2016).
8. Watson, R.N.M. et al. CHERI: A hybrid capability-system architecture for scalable software compartmentalization. In *Proceedings of the 37th IEEE Symposium on Security and Privacy* (San Jose, CA, May 18–20, 2015).

Peter G. Neumann (neumann@csl.sri.com) is Senior Principal Scientist in the Computer Science Lab at SRI International, and moderator of the ACM Risks Forum.