Insider Threat

James Bore

Abstract This chapter discusses the threat arising from within the organisation, whether from negligence, malice, or exploitation by an external party. The trusted insider is one of the greatest challenges facing organisations today. The analysis considers the balance to be struck between allowing insiders access and privileges to show trust and increase productivity, and securing that access at the cost of good will and with an increased risk of workarounds being found, placing vulnerabilities at the heart of an organisation's policies and processes. The tactics of social engineering and exploitation of human psychology to compromise or completely bypass technical and procedural security measures are considered, along with the effectiveness of training and difficulties of raising cultural awareness of security on a long term basis in a rapidly changing technological landscape.

Keywords Authentication · Autonomous devices · Impersonation · Insider threat · Smart devices · Trust · Social engineering

1 Insider Threat

A common misconception is that cyber security incidents, particularly data breaches, are the result of sophisticated attackers who have found a way through a company's defences. Even where there is good awareness of the threat that insiders can pose, an assumption is often made that it is the malicious insiders who are at fault and seek to deliberately cause harm to the organisation. Fortunately these misconceptions are quick to disprove. Data gathered by the Information Commissioner's Office (ICO) clearly shows that the overwhelming majority of data breaches occur due to mistakes made by trusted parties, whether they are

J. Bore (✉)
Independent Researcher, London, UK
e-mail: james@coffeefueled.org

H. Jahankhani et al. (eds.), *Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity*, Advanced Sciences and Technologies for Security Applications, https://doi.org/10.1007/978-3-030-35746-7_19

configuration errors in systems, messages and files sent to the wrong party, or simply misplacing paperwork [15].

The ICO data discussed above only provides detail on incidents which have been both detected and reported. Given the common desire for individuals to cover up or deny mistakes, combined with the under reporting of cyber security incidents suggested by [16], it is likely that there are significantly more incidents caused by insiders which are never reported even within the organisations affected.

As the potential for interconnectivity and greater communication options opens up, so too do the opportunities for simple mistakes or omissions to cause incidents. Greater data storage, more access to sensitive information and systems, and the increase in automation with reduced manual oversight means that a simple mistake can quickly magnify from a minor incident to one with severe impacts on nations, or even internationally. Recent cloud outages caused by administrator configuration mistakes illustrate this well, a Border Gateway Protocol (BGP) misconfiguration by a Verizon customer, along with a failure of filtering by Verizon, affected several thousand networks and tens of thousands of individual IP addresses. Packet losses caused damaging issues for several major providers, including an estimated failure of 15% of traffic to the service provider Cloudflare along with impacts to Amazon and Linode. All the damage was due to a relatively simple configuration error made purely through negligence.

All of this is without considering the potential of genuinely malicious insiders, or of external attackers seeking to exploit trusted parties to use as a proxy into organisations or systems. While this is much less prevalent in terms of causing incidents, the damage from a deliberate, targeted attack by either a malicious insider or one who is being manipulated by an external attacker is significantly greater. An e-mail sent to the wrong recipient is much less likely to reveal sensitive information to a person equipped and motivated to make use of it. With a targeted attack, the attacker knows exactly which information they are after, or which system they want to target, and will have prepared to exploit it effectively beforehand.

Even where negligence on the part of a trusted party does not directly cause an incident, it can contribute greatly to opportunities for malicious attackers. Both the well-publicised British Airways and Equifax breaches were due to errors made by trusted parties. In the case of Equifax a simple oversight in vulnerability management led to the compromise of records covering nearly half of the United States population, along with millions of records related to inhabitants of the United Kingdom and Canada. While this is generally seen as a breach by an outside attacker, the role of the insider in enabling the attack cannot be overlooked.

When we add new technologies under development to this picture, we can clearly see a path where the insider threat will continue into the future. The development of technologies such as Deepfakes [22] potentially allows attackers to exploit a trusted insider's position without any direct interaction with the victim. With the increase of remote working and newer methods of communication such as holographic communications the potential for attacks enabled by simulation and impersonation of trusted insiders becomes clear. As the Internet of Things (IoT) expands there are also questions about autonomous and smart devices which require human oversight

– an illustrative example is autonomous cars, where even the definition of the insider becomes challenging.

For discussion purposes we will define the insider as anyone or anything fulfilling a trusted role. This allows the inclusion of impersonators who make use of presumed insider status in order to exploit an organisation, those who cause or enable incidents through negligence, and implicitly trusted devices within systems. The insider threat can then be defined as any use of a permission arising from a trusted role which goes against the intended purposes.

While there are certain basic common defences for insider threats arising from negligence, malice, and impersonation each have certain attributes that demand a more focused examination for effective mitigation. There are also, as always, aspects of the insider threat which fall into grey areas such as shadow Information Technology (IT)[1] which does not fall neatly into either the negligence or malice category, but would perhaps be better thought of as misapplied good intentions leading to increased risk.

1.1 Preventing Negligence

Negligence is the most common and insidious form of insider threat, difficult to predict and nearly impossible to defend against comletely. As the poet, Alexander Pope, stated "to err is human". There is no evidence that this is going to change in the future, and indeed as the technological landscape expands our opportunities to get things wrong it is much more likely to increase. Given that, at a fundamental level, all insider threat can be reduced to the idea that people use a system or process for a purpose or in a way that is not intended, the methods we can use to treat negligence are universally applicable across any form of insider threat.

The simplest applicable defence, regardless of technology level, person, or role, is the principle of least privilege or principle of least authority. Comprehensively enforced, the least privilege principle prevents individuals or devices from being able to do anything with a system that they are not intended to be able to do. Unfortunately this is not a perfect control to prevent insider threat, whether malicious or negligent, as it is abuse of the permissions that they are given that causes the threat to manifest. However, placing strict limits on the functions available at least makes the scope of the threat definable for a particular role, and allows other appropriate controls to be put in place.

One of the major challenges of this approach is that overly strict controls, without effective training and communication, can enhance rather than reduce insider threat as individuals find alternative paths to complete work when systems are seen as

[1] The use of solutions or systems by individuals or groups within an organisation outside the oversight of a governing IT or cyber security function, more common and often seen as inevitable within large organisations.

awkward to use. Whether the controls are policy and procedure based, or enforced through technology, this will apply – causing procedures to be ignored and the rise of shadow IT. Worse yet, the deployment of shadow IT outside the oversight of departments responsible for governing security is sometimes seen as a positive by the senior leadership of an organisation for knowledge sharing and collaboration [20], as the rules and standards put in place by security teams are still occasionally seen as overly strict, and unjustified.

Even where the principle of least privilege has been thoroughly applied, assuring that no single entity has any more rights than required to fulfil their role, there is a second requirement that is necessary. Ensuring that no single person can perfect potentially abusive actions requires separation of duties between different roles – a typical example is that the person who makes a payment should require authorisation by another role. In a small organisation, the principle of least privilege may not prevent a person from authorising their own payments, and so we must apply separation of duties.

Separation of duties entails ensuring that no one person or other entity has permissions to both enact and authorise any potentially harmful action. Even in small organisations where persons may need to both authorise and enact payments this can be implemented through the mechanism of ensuring that no person can authorise their own payments. Despite being simple in concept, there are many difficulties in implication and often enforcement will become a case of monitoring for breaches rather than a technological control preventing the actions.

Either malice, or willful negligence, can cause separation of duties to be circumvented by persons working together in undesirable way. Preventing collusion is more challenging than preventing simple abuse of privileges, and any organisation must make a decision on how they wish to strike this balance given their risk appetite and the potential for abuse. Again we return to the issue of trust, as an organisation may decide to invest higher trust in its employees, allowing for greater and more efficient productivity with the increased risk of negligent or malicious activity causing greater harm.

Even where collusion may be guarded against through layers of independent approvals and independent audit, one similar area of insider threat is where conflicts of interest arise. A senior employee with a vested interest in a particular supplier may have the authority to direct purchases their way, which can cause not only issues with fraud and corruption but also weaken security controls since full due diligence and effective evaluation of services and products may be bypassed. Particularly for senior decision makers guarding against conflicts of interest is essential to reducing insider threat.

Without effective communication and cultural change, security controls applied to individuals may not only be ineffective (as they are worked around), they may actively encourage the insider threat to arise. The balance of encouraging insiders to feel trusted and responsible rather than restricted, not hindering productivity, and yet still installing effective security controls whether through organisational or technological means is challenging. As it depends almost entirely on human factors, each organisation must consider carefully the approach appropriate both to their level of accepted risk and the culture which exists. Where there is a greater sense of responsibility among individuals, greater controls can be placed without significantly increasing the chance of them either being worked around or motivating insiders to become disgruntled and act against the organisation.

Implementing the form of cultural change needed to reduce insider threat is a challenge, depending largely on the specific organisation involved as well as the wider surrounding culture or cultures. This is especially difficult for multinational organisations, as different national cultures have a significant impact on the perception of cyber security risk and the degree of responsibility individuals feel towards their employer. It is clear that the simple approach of a general cyber security awareness campaign is often ineffective [3] and a more targeted approach is required. Any such approach must take into account the different audiences within an organisation, aiming to foster feelings of ownership as well as understanding of the security issues.

The simplest way to defend against negligence as an insider threat is to apply automation wherever possible. The debate about the positives and negatives of automation aside from in terms of security is still ongoing and heated, and there is no need ot delve into it here. In terms of security and especially the insider threat, however, automation is a way to prevent simple human error in a process from causing damage. As a bonus, automation also makes the malicious insider threat, whether through an active insider or manipulation by an outsider, significantly more challenging. Fewer human entry points into a system allows for more thorough defense of those that remain.

In essence there are five broad types of overlapping human-enabled insider threat that can apply to an organisation.

Willfully ignorant There are those who are willfully ignorant, refusing to take part in training or awareness programs and thus become ready vectors for a malicious attacker, or for the negligence and collusion possibilities previously discussed. An effective audit of training and awareness programs will highlight these over the longer term, and an organisation must act to address them. This class overlaps heavily with the next in that those who have fallen foul to phishing attacks previously, and refuse training, are more likely to fall foul to repeated attacks.

Negligent As mentioned, negligence is the most common form of insider threat as well as the most challenging to resolve. While it is possible to carry out training and awareness, it is a simple fact that even those who normally exhibit highly secure behaviours can and will make mistakes. Even when reviews and approvals are put in place, it is possible that highly trained reviewers will miss things, meaning negligence can never be fully resolved as a threat. Negligence is also the most common vector for external attacks to succeed, through targeted social engineering attacks, system misconfigurations, or other mechanisms.

Collusion While this area does overlap somewhat with willful ignorance and negligence, it tends to appear more in combination with malice. Collusion is most commonly the domain of fraud and similar activities, though theft of intellectual property can also occur. Collusion may also not purely involve the insider, but may be an active deliberate cooperation with an outside actor.

Resentment A previously trustworthy employee who feels mistreated can be the most damaging threat to an organisation. One of the most dramatic examples was heard in [24], where a network administrator, Terry Childs, after a dispute about having lied during his application process, was found to have seized sole control over the FibreWAN networking infrastructure for the City and County of San Francisco, and refused to disclose the passwords required for access. While the passwords were recovered in a short time, between the 9th to 21st of July the city government had no control over the FibreWAN infrastructure and, due to the mode that the configurations were stored, were unable to risk any power disruptions which may have affected the network.

Persistent insiders While a resentful insider is likely to cause the greatest single incident of damage to an organisation, a persistent insider is a much greater long term threat. Usually they are employees seeking to use company resources for a supplemental income, and often have the access and authority to prevent discovery for a significant amount of time. It is only as the damages they cause increase over time that detection becomes more likely.

1.2 Prediction and Detection

Sanzgiri and Dasgupta [28] provides a useful taxonomy to look at the methods available to both predict and detect insider threat activities. The methods range from technological controls such as the well established Role Based Access Control (RBAC) model and deployment of decoy or honeypot[2] files, to predictive methods based on psychological and behavioural factors drawn from monitoring of employees. The addition of threat modeling techniques [14] allows for prioritisation in deploying these detection methods, as well as giving guidance on where and which controls to prevent incidents should be deployed.

Behavioural analysis, in particular anomaly detection, is currently a popular way to detect potential insider threats. Network traffic, application activities, use of permissions, working hours, and various other factors can be and are considered as inputs to a model applicable to a particular role. The baseline is normally built up over several months before alerting is switched on, meaning that with a sufficient sample size it is unlikely that consistent anomalous behaviour on the part of one individual will affect the model for a role. Of course, there are significant ethical concerns with these models and, while they are effective at detecting deviancies from normal practices, there are multiple explanations available. It is important that

[2] Files or information designed to appeal to a malicious party, often with dummy sensitive information. In more advanced implementations the files may have dynamically generated watermarking such that individuals will be linked to unique files, making investigation a simple process.

any such behavioural analysis is not only carried out with careful consideration of the ethical issues involved with observing employees, but also that any alerts which arise are handled with consideration and care since mishandling of an anomaly arising due to causes unrelated to insider threat are more likely to make such a threat manifest than prevent it.

To give examples of how behavioural analysis can be applied, and the associated difficulties, we will look at working patterns. If we take the example of an employee whose working hours are normally between 9am and 5pm, with some variance for travel disruptions or staying late to finish projects, a baseline can quickly be established. If there are then consistent anomalies outside of that baseline, let us say working until 8pm several days in a row when very few others are in the office, gives an indication that something has changed. Various causes could be put forwards, maybe a particularly complex project the employee is working on, a change in financial circumstances meaning overtime pay is required, a personal circumstance that means they want to stay longer at the office, or possibly something more sinister such as collecting sensitive information for sale or distribution. This is where the ethical concerns, and the need to treat the situation carefully, come into play. With any of the causes which are personal, trying to restrict access to information or starting any overt investigation could easily be perceived as persecution by the employee causing an otherwise trustworthy employee to become less so. Of course, the other causes could also be a trigger for a less malicious insider threat to manifest, so not doing anything to mitigate the situation is also not an option.

All of the behavioural analysis methods for detection converge in the concept of the digital twin, and it seems likely that existing technologies for insider threat detection will begin to move in this direction – if they have not already done so. Providing behaviour profiles for individuals in the form of a digital twin[3] which can be subjected to much more thorough analysis than the individual otherwise would does have significant potential for improving the prediction of insider threat, though the ethical issues in such deep profiling by employers, including the incorporation of information from social media platforms and various other sources, are difficult to consider. At the least it is likely that digital twin-based predictive technologies will be reserved, initially, for highly sensitive organisations where the requirement for security outweighs the right to privacy for individuals. This already occurs, to some degree, with the highest levels of security screening.

Authenticity analysis with machine learning algorithms assisting human participants will become increasingly vital to insider threat detection and prevention as emerging technologies, such as holographic, Augmented Reality (AR),[4] and

[3] A synthetic behavioural model of a person or other entity used to predict behaviour, whether predicting drug interactions in medical research or purchasing patterns in online shopping.

[4] Augmented Reality is use the use of technology to enhance the real world using technologies such as digital overlays to provide contextual information, or insert artificial digitally generated 3D objects into a real-time landscape.

Augmented Human (AH)[5] communications for collaboration become well established. Given the existence of deceptive technologies which will be discussed in the next section, and the pace of development, it will become essential to make use of detective technologies in any sensitive remote communications, whether video, audio, or holographic, to assist human participants in guaging authenticity [6]. This is already occurring at a basic level with e-mail communications, with many systems now raising notifications that a recipient is external, or that an e-mail is suspicious. Given the profitability of deception-based cybercrime exploiting insider threat it is not likely that bad actors will ignore the potential to use new technologies to support their attacks, and continue to manipulate insiders [25].

1.3 The Outside Insider

Insider threat does not always relate directly to human employees. The key requirement is that an attacker somehow gains access as a trusted individual, and there are many examples of how this can occur without ever directly engaging with an organisation. One of the more recently named techniques which illustrates this effectively is warshipping [13]. Named as a portmanteau of shipping and wardriving, which involves individuals travelling around searching for wireless network signals in order to gain access, warshipping exploits the delivery network of companies. Even where a company may take great precautions to prevent their wireless network from being publicly accessible, warshipping bypasses all of the physical protections and provides an attacker with an evil access point within the organisation's premises.

Conceptually warshipping is a very simple attack: the attacker will build a small, battery powered computer with both wireless and mobile connectivity options. The mobile connectivity is used to remotely control the device, to then compromise any wireless networks detected when it has reached its destination. Location is monitored using mobile connectivity to determine which cell the device resides in, or GPS. Once it has arrived the device will continue working for days or weeks depending on the battery available. Known exploits against wireless networks are used to gain access, allowing the attacker to carry out Man-in-the-Middle (MitM) attacks or other compromises.

Naturally warshipping, while recently named, is not a new technique though its prevalence has not been studied in depth. Of course, often there is not even a need to go to the lengths of paying for shipping, as often an outsider can pose as a visitor to an organisation and gain access to public areas. This gives easy access to any organisational wireless networks in the building, and in some cases direct network access where port management is not correctly implemented. Even where port management is properly managed an outsider who can gain access to secure

[5] Using wearable or implanted technology to enhance the capabilities of a human.

areas, whether by pretexting as an employee, visitor, courier, supplier, or another role, can make use of a wide variety of tools to manifest an insider threat with only limited contact with employees. A well known attack which exploits the insider, with minimal risk to an attacker, is simply dropping USB drives in an area near an organisation's premises, ideally marked with some interesting and relevant label to exploit the curiousity of employees. Once inserted the device carries a payload of malware which allows the attacker to act in the role of the manipulated insider on the networks and systems.

For better-resourced attackers rather than travelling in person, or paying for shipping, drones may be used to the same effect. Some of these are autonomous, compromising wireless networks and building networks of insider bots on the fly [26].

Smart devices often used with voice activated smart home systems, or with smart phone control, such as lights, kettles, aquarium control systems, pet feeders, and similar convenient items. These are often provided with trusted access to home and office networks, and the lack of any comprehensive security standard governing their design and manufacture makes them ideal entry or manipulation points for an attacker. Since all of these devices will hold, at least, the access keys to a network the insider threat is clear. Of course, the threat profile of some devices is significantly higher than others, especially when safety features are implemented in manipulable code rather than mechanical. It is not a great stretch to think of a kettle with temperature regulation governed by a thermostat controlled in software, to allow for custom temperature adjustments for the perfect cup of tea. Unfortunately such a device, without a mechanically-implemented safety cut-out to prevent overheating could easily be the cause of a fire, moving the insider threat from a threat against information to one in the physical realm.

Plenty of high-profile examples covering why these devices should not be trusted, and should at best be isolated to dedicated, quarantined networks are available. One of the most popular tales is of a casino, which will remain unnamed, which serves as an example of insider threat both through a smart device, and through the actions of a customer or visitor as an insider. Originally discovered by Darktrace, the threat manifested when a visitor to the casino connected to the fish tank control system. The control system was connected to the internet in order to regulate temperature and feed the fish when needed, but was also connected to the casino's internal network. Once on the network the attacker's located the high-roller database, and exfiltrated it through the thermostat to cloud storage [30]. The attack on the casino also highlights the importance of applying least privilege principles to devices as well as individuals: there is no need for a thermostat to be able to transmit data to cloud storage, or even run a fully functional computing system. The scenario could easily have been prevented by limiting the device to only accepting and sending temperature data, with anything else being dropped automatically.

For a more literal view of the insider threat, implantation of medical devices has saved many lives. As technology advanced, developments have meant that controlling, adjusting, or updating medical devices no longer requires invasive surgery as it would before, reducing the immediate risk to patients. Unfortunately this does raise other risks instead, as demonstrated by [27]. Butts and Rios discovered vulnerabilities in various medical devices, with the most dramatic being the control system for a pacemaker. As the AH becomes more common, whether with life-saving medical devices, enhancements, or simply convenience devices such as implanted chips it is vital that this area of insider threat, sitting inside the target's own body, receive the attention it deserves. The ease and simplicity of other attack vectors, along with the morality involved in compromising a medical device, may help to keep the threat reduced, but it is not hard to envision a form of disturbing ransomware deliberately targeting devices if they are not adequately protection.

Until the ongoing rise in smart devices is matched by increased security awareness and understanding in manufacturers, this is an area of insider threat that will continue to grow.

Combining two of these vectors, the SkyNET botmaster system and similar networks can be used in order to find and compromise smart IoT devices in addition to wireless networks. The first widely popularised system making use of this strategy was developed by the security firm Praetorian in 2015, and is continually developed to this day. While their demonstration does not attempt to compromise discovered smart devices, it provides an effective picture of the attack surface for those whose attempts are more malicious and was designed to highlight this potential. A similar service, without the use of scouting drones, is provided by the search engine Shodan and searches for smart devices directly attached or routed to the public internet. All of these devices are a potential vector for insider threat to manifest, as they are almost always considered trusted participants in their host networks.

1.3.1 Insecurity by Design

Issues with the security of smart or other computing devices are not always due to failings in design. Deliberately introduced vulnerabilities, or backdoors, are a major concern for organisations from the scale of small businesses to national governments, and even individuals. Whether introduced for the convenience of developers and not removed after the development process is complete, or maliciously inserted into the design process by a hostile actor, the threat posed by backdoors in hardware or software is increasing.

Software backdoors can be treated after discovery in the same way as any other vulnerability, with patching or software updates. Unfortunately many backdoors have a hardware component that makes their removal range from challenging to ouright impossible. One of the earliest well-documented cases occured in 2008 with the leak of an Federal Bureau of Investigation (FBI) presentation detailing the investigation into counterfeit Cisco components which had been installed across the United States, including in sensitive military and government sites. While Cisco claimed the counterfeit equipment was for profit reasons rather than corporate or governmental espionage, the FBI considered the threat of backdoors being included severe enough to describe it as a ‘critical infrastructure threat’. Worsening the situation were claims that Cisco's approved vendors had also purchased and resold counterfeit equipment.

The potential of a compromised trusted supply chain leading to the installation of malicious equipment in key systems is not taken lightly, and is not simple to resolve. Full inspection of every piece of equipment, down to circuit board and silicon wafer level, would be required to remove any trust requirements given the sheer complexity of modern systems. Simultaneously that level of deep inspection is not merely impractical, but impossible with current resources and tools. Trust is required for any such chain to function effectively, and as with individuals the more controls and checks are applied the more impact there will be on productivity. Indeed, as with employees it is entirely possible that excessive arduous checks will simply lead a supplier to decide to cut their losses and walk away. When this is a matter of specialist hardware, the receiving party may simply not have an alternative but to allow the checks to be bypassed.

The Department of Defense (DoD) takes the situation seriously [8], as do many other national governments and organisations. Unfortunately given the complexity of systems, as the report states, it is extremely challenging to very some of the claims made about supplier compromise, or malicious suppliers, and even more so to separate the claims and subsequent actions from other political or economic motivations. When claims of security risks from supposedly trusted sources have been shown to serve an ulterior motive, even our own sources of security intelligence can become an incidental or malicious insider threat. Take as an example a claim that a major equipment manufacturer is in some way compromised. If this claim is made by a significant and trusted authority, such as an intelligence agency, it is likely to cause equipment to be removed, supply agreements to be broken, and similar. If this claim was made not on the basis of genuine security concerns, but for national economic interests, then at best it is a waste of effort and funds, while at worst it could lead to further threats arising from the use of substandard systems.

1.3.2 Inside Voice

It is not always necessary for an insider threat to be overtly hostile as in the case of warshipping. More and more homes are now adding some automated capability through the use of voice activated or remotely controlled smart home devices. While there is no denying the convenience of these devices, they do have a substantial and often overloooked threat profile. The instances of children, or in one notable case a pet parrot [4], purchasing substantial amounts of food, treats, or other undesired items (at least undesirable to the person paying for the purchase) are well recorded.

In 2014 various owners of Microsoft's, at the time, new XBox One console began complaining of the devices activating in response to a new television advert which began with the words "XBox On". At the time those affected viewed the issue as a minor nuisance, with the greatest damage being from those who would find their XBox activation would switch the inputs of their televisions [17]. In April 2017 Burger King took things a step further, by using their television advert to deliberately query the Google Home smart device with the question “OK, Google, what is the Whopper burger?”. The advert, and resulting irritation of viewers, led to a brief edit war over the Whopper wikipedia entry before Google stepped in to disable the voice query [31]. A less deliberate instance occurred in early 2018, when an advert for Purina cat food contained the words “Alexa, reorder Purina cat food”. While that instance did not result in a confirmed order, it followed a spate of similar instances which did and caused vendors of smart home devices to add layers of confirmation before orders would complete, along with blocking their own adverts (and any others they were notified of) from triggering the voice recognition [1]. Given the sheer number of voice activated devices now carried and trusted by users, the potential for insider threat through voice recognition is obvious.

More concerning is the fact that much of the voice recognition processing for these devices is carried out on cloud-based systems, and it has come to light that to assist these machine learning models human contractors have been provided with audio recordings of conversations believed to be private in order to improve recognition success [7]. Any voice activated system should now be considered suspect and a potential insider threat, no matter how much convenience they add. Voice recordings and transcripts can be requested under various privacy legislation, and in previous cases a request has resulted in the wrong individual’s data being sent, though it can be hoped that this was a one-off case and the lessons learned were quickly implemented [2]. The misinterpretation of a conversation as voice commands is not so easily addressed, and even has a higher potential for damage since it may result in private conversations being shared with friends and family [18].

1.3.3 When Security Is a Threat

Even where devices and infrastructure are secured, much of the time they are now secured by various third party managed services. Alternatively many companies, recognising that the expertise required to provide cyber security capabilities is a specialist area, will delegate responsibility for their security to third party providers who take on not only the administration and configuration, but the hosting of their tools.

This approach has definite upsides as economies of scale kick in, particularly with shared providers. The problem comes in when those shared services also raise potential threats themselves. A Managed Security Service Provider (MSSP), or even Managed Service Provider (MSP), can certainly provide security expertise which may not be available in house, but they simultaneously increase the threat landscape for their customers. By becoming a single trusted provider to multiple customers they make themselves a valuable target for attackers or any form of malicious insider, the trusted access to customer systems providing a single point of entry for multiple further targets. Given the size of many MSPs and MSSPs organisations their structure and infrastructure have significantly increased complexity, which opens up many more opportunities for anyone with both access and malice in mind.

Due to this companies need to consider carefully what access a third party provider should have to their systems, and again apply principles of least privilege and separation of duties. Due diligence to ensure that a provider is secure themselves is also essential. A threat actor targeting an MSP is likely to be much more sophisticated and better resourced than one targeting a single small to medium enterprise, and while security by obscurity is very negatively perceived in modern times, a lower profile against motivated and well-resourced attackers is definitely of benefit.

1.4 The Future Insider

As technology advances, our systems grow in complexity and become ever more interconnected. This interconnection between technologies and people, and the increasing rise of AH through wearable and companion devices blurring the line between users and participants in networks, means that the threat landscape is constantly growing at the same time as more and more participants and devices are becoming insiders. A city-wide public wireless network, as an example, means that every person within that city with any wireless device becomes a potential insider and target. To add to the landscape we must also consider non-human network participants such as more advanced, autonomous smart devices, and even the overarching machine learning programs that govern these systems as potentials for insider threat.

On top of these, the growth of AR and communications technologies to permit long distance and asynchronous communications, whether through telephony, email, or collaboration tools, means that the social network of insiders continues to expand alongside the technological underpinnings of society. When physical presence is no longer necessary, establishing the authenticity of a participant becomes exponentially more challenging. Even with technologies to allow for visual communications, emerging systems make it harder to establish whether the parties at the other end of a long distance connection should be treated as trusted insiders, or quarantined from a system. Finally, such long distance communications remove one of the most effective controls on insider threat – face to face communications allow for a much more effective assessment of an individual's behaviour and attitudes in most cases than a remote call, and many insider threats have been prevented by the simple expedient of colleagues in a workplace noticing that someone is behaving oddly, unusually angry, or seems tired and has been making mistakes due to personal circumstances.

1.4.1 Smart Transport and Smart Cities

Deserving of separate consideration to simple smart appliances, as almost the ur-example of a futuristic cyber security threat in both media and research, are autonomous vehicles [29]. A popular scene in modern films looking towards the future is a car chase (sometimes with flying cars) where between one to all of the vehicles involved are compromised by an attacker, acting as hostile agents in a trusted network. While such scenes once seemed outlandish, and are usually presented as a hostile outside attacker, the use of trusted, open networks to carry out this form of attack means we can consider it as an insider threat.

Another example, older even than the autonomous car compromise, is the attacker who targets a smart city's networked traffic management system to similar effect. Vehicles under the control of humans are attacked through causing traffic control systems to show contradictory signals. While again this is arguably an attack by a hostile outsider, the threat itself manifests through the misbehaviour of a trusted network system. Such an attack would be ineffective if the system were not trusted and relied upon by all participants.

At this point even loose definition of insider with which we began this chapter becomes a challenge. Is a car which is misled through bad radar information or control codes a threat to the driver? Is a driver through inattention and poorly-judged manual override the threat to the car and autonomous road systems? Is a pedestrian, or third party driver a potential insider threat to an autonomous road network in a smart city, or vice versa? In these scenarios, are one, both, or neither insider threats or external threats? There is no clear answer to these questions, and any answer given depends entirely upon individual perspective and the context of the discussion.

For simplicity we will say for now that in a smart city, autonomous travel system, or similar shared publicly accessible network, all users and authorised participants can be considered as insiders, and potentially give rise to insider threat. The open nature of such systems increases the number of insiders and simultaneously decreases the amount of implicit trust they can be safely granted, since unlike employees or contractors there are fewer options for instilling an effective security culture, and the greater variety of participants leads to a much higher likelihood of undetected disgruntled or even malicious insiders.

1.4.2 The Unpredictable

Big data, machine learning, and digital evolution for problem solving are popular approaches as our technological landscape becomes too large and complex for individual, or teams of, humans to comprehend in any realistic timescale. Machine learning is ever more sought after as an approach to analyse events in pursuit of anomalies that might indicate security incidents in large computing networks, in addition to being applied to much more diverse problems such as landing planes, planning road layouts, managing traffic, profiling customers for targeted advertising, monitoring human users for abnormal behaviour, carrying out medical analyses, and many other purposes. As these models and methods are developed and applied, the outputs are given a huge amount of innate trust, in some instances more even than the analysis that might instead be produced by a team of humans.

As machine learning models become trusted participants and sources of intelligence, we must be on the watch for insider threat. This is not because our current machine learning models are likely to become self-aware and take over the world, but because they are unpredictable and, in many cases, comprehensible only after mistakes are realised. In many cases these systems are tested in isolated, simulated worlds, set with a certain collection of goals before being deployed to any real-world function. A few of the narratives described by [19] indicate the dangers of implementing any system which has not been fully understood by its creators – when we are discovering exploitable vulnerabilities in light bulbs, there is a very valid cause for concern in implementing algorithms which are not comprehensible to their creators and the risks posed must be considered carefully.

It is inevitable that, as these systems are implemented, a proportion will have been set up with incomplete initial goals, leading to such situations as the algorithm to land a plane instead bringing it down with enough force to overload the sensors, fortunately in simulation rather than practice. While this is an extreme example, where consequences are less dramatic there is a lower chance of detection during development and testing, meaning that decisions will be taken based on incorrect analysis. With the complex environment we experience today it is inevitable that machine learning algorithms utilising big data collection will be implemented in order to solve challenges, and we must be ready to deal with this new vector for insider threat to manifest.

This does not account for people misusing the systems implemented using greater information to their own benefit. Multiplayer online gaming illustrates that participants in a system, in pursuit of competition, will learn to exploit flaws in the system to their own advantage and the detriment of others. Where such systems are implemented in a wider environment in order to govern rules of behaviour it is likely that a number with advanced knowledge of the system may do exactly the same, learning to feed carefully determined misinformation to the system in order to benefit themselves at the expense of others. Indeed, this happens currently in many systems of bureaucracy where fraud is a serious concern. The attempted impeachment in 2018 of the Supreme Court of Appeals of West Virginia serves to highlight the effectiveness of privileged actors abusing their positions, as well as the difficulties of detecting such behaviour and mitigating the damages after the fact.

1.4.3 Consequences of Mitigation

As with most insider threats, the most effective mitigations for these large scale environments with a broad population of insiders can be reduced down to two options. One is to reduce the amount of trust in the insider, applying greater and greater automation to these systems with fewer points of interaction and reduced control for participants. This reduces the potential attack surface for a disgruntled insider, but does have consequences in terms of human autonomy in such environment by reducing options and applying ever-stricter controls in pursuit of a secure environment. Alternative there is the option of attempting to inculcate a security-focused and highly responsible culture using the autonomous environments themselves. The measurement of trust and scaling of potential interactions based on rule-following is undoubtedly a security measure, but the human cost involved in terms of privacy loss, reduction of rights, and potential for abuse is not something to be dismissed lightly.

China has been implementing and developing one such cultural regulation system since 2014, and the ethical debate across the rest of the world is no less heated years later than when the system was first proposed [5]. Concerns can also be raised over the increased importance such a system places on digital twins over the behaviour of people when not observed by the system, and the potential for calculated attacks against the behavioural profile of participating citizens through a co-ordinated misinformation campaign are of serious concern. Considering the real life impact which can occur as a result of co-ordinated social media shaming or harassment campaigns already, the possibility for such attacks leading consequences enforced directly by a social credit score supported by governments is truly terrifying. All such incidents are a form of insider threat, as those targeted by and co-ordinating the campaign must be authorised participants in the system.

Equally the disruption that could be caused to a state relying on a governing social capital system, tied into all automated systems and with oversight on booking and other commercial systems, by a single bad actor with high level privileges is almost staggering. Whether against an individual, or against a large group, an administrator would potentially be capable of bringing an entire society almost to a halt by maliciously adjusting the privilege thresholds for citizens. In such a society, those who directly control the system and its design have the greatest potential to cause damage to the society as a whole.

1.4.4 Impostor

Currently spoofing and impersonation is largely limited to changing source email addresses, display names, or phone numbers. SMS spoofing in particular is becoming prevalent as banks and other service providers switch towards using One Time Passwords (OTPs)) sent via mobile networks. An attacker using a spoofed number can then impersonate the service provider, and often will ask for the OTP sent to be forwarded on. Similarly a phone call from a spoofed number can lead to the same result, circumventing the security normally provided by Multi-Factor Authentication (MFA). All spoofing and impersonation attack techniques can fall under insider threat, as they attempt to exploit the implicit trust placed in insiders.

The use of deceptive technologies such as Deepfakes [11] in order to spread misinformation is already well documented. As the technology develops and is combined with face tracking and recreation systems [33] the potential for use as a tool for malicious insider threats, or indeed carrying out an insider threat without requiring any participation on the part of an insider, becomes clear. Given the data and behavioural profiling that is considered a standard, it is entirely possible to conceive of a dedicated, motivated attacker constructing the profile of an insider based on personal data leakage in order to carry out an attack. While this may seem outlandish, it is already one of the most effective techniques in the social engineering arsenal, with impersonation being one of the most successful an profitable mechanisms to extract funds from organisations [21] with the use of nothing more sophisticated than changing e-mail display names.

Arguments have been made that an impostor is not truly an insider threat. To counter this it is clear they are acting as an insider and looking to exploit the implicit trust involved to their own ends. The only additional precautions to be taken against an impostor threat, beyond those that should be taken against an insider threat, are methods to verify identity and to put insiders on guard against potential impostors. Otherwise, all other controls placed for a true insider threat are equally effective against an impostor.

It is obvious that this particular insider threat is going to increase with time. Currently many uses are for entertainment, though there are already instances where Deepfake systems have been used to generake fake pornography of both celebrities and individuals, whether for blackmail, exploitation, or revenge purposes. Currently it is usually possible to establish the authenticity of such videos, but this technology is still in the earliest stages and the possibility of a world where we cannot reliably trust any video recording or communication as being authentic is a genuine concern. New authenticity mechanisms will be required, whether they will involve a resurgence in physical meetings or other security mechanisms such as assuring secure connections through the use of security passphrases which are provided by some secure communications applications.

Several sites already exist allowing reading aloud of any text in the voice of a particular individual, and faked videos of personalities such as Barack Obama have circulated since 2017 [10]. While it may take some time before this deceptive technology expands to include holographic communications through AR, or to insert false avatars into Virtual Reality (VR) or even classical video conferencing environments, the tools are all available with voice recognition providing a realtime input mechanism, facial tracking allowing for ever more realistic imitation of an artificial face in terms of emotion, lip-syncing, and general expression, and Deepfakes applied to create the artificial avatar and replace the impostor's voice.

We are likely heading towards a future where the current cyber security arms race between defenders and attackers will expand along yet another front, with authenticity of communications previously trusted implictly outside of highly sensitive environments become an ever-greater concern. While research is rapidly taking place into various methods to defeat deceptive technologies, many of the answers depend on particular artifacts which are a result of the immaturity of the technology and will be rapidly addressed as its usefulness as a tool of attack becomes apparent [22, 23, 32].

As an aside, it is interesting to note that these sophisticated attacks on authenticity do not apply only to information and communications, but have crossed into the physical world. Especially in the art world, where machine learning algorithms are often used to establish authenticity, the threat posed by algorithms which can produce 'replicas' which appear authentic to an artist's style is clear [12]. With more direct implications on general security, the potential for counterfeit documentation of all kinds can only cause trust to be lost in physical copies, requiring more dependence on computational authenticity mechanisms such as cryptographic signing authorities as are now provided by the Estonian government [9].

1.5 Overview

We have seen that the insider threat is a broad subject, with many different attack vectors ranging from simple innocent negligence to a well-resourced and motivated attacker posing as an insider. As our technological systems develop opportunities for these threats to manifest continue to increase and develop, providing new vectors for attack which leverage brand new technologies and quickly adopting new capabilities to serve their goals.

New communications channels will require new methods of working to ensure authenticity, where we have not achieved this even with well established channels such as email. Ever more complex systems give rise to more potential for damage to occur through both negligence and malice. The increase in technologies such as AR and particularly the rise of AH, along with smart devices and increasing dependencies on interconnected systems mean that threats from insiders, including participants in systems, can manifest in new and unpredictable ways.

Whether the insider threat is a true insider, an imposter, or the manipulation or exploitation of an insider to achieve an attacker's ends, certain founding principles can help to mitigate the threat. Building systems, where possible, or re-engineering systems to apply least privilege principles and enforce separation of duties will help to prevent insider threat from manifesting for an organisation, while instilling a culture of security awareness and responsibility where insiders can be trusted will reduce the possibilities for manipulated insiders to cause damage to an organisation.

Regardless of the measures taken to prevent insider threat from manifesting, the nature of people and the requirements for trust in order to achieve anything as any organisation means that the threat and its accompanying risks will always be present. Properly ascertaining and addressing these risks can help to minimise them, but they can never be completely eradicated from any organisation and constant monitoring is required to ensure that when they arise they can be dealt with in a timely and effective manner.

References

1. Ad That Fooled Amazon Device Cleared (2018) In: BBC news. Business. https://www.bbc.com/news/business-43044693 (visited on 27 Aug 2019)
2. Alexa User Accesses Stranger's Chats (2018) In: BBC news. Technology. https://www.bbc.com/news/technology-46637427 (visited on 27 Aug 2019)

3. Bada M, Sasse AM, Nurse JRC (2019) Cyber security awareness campaigns: why do they fail to change behaviour? arXiv: 1901.02672 [cs]. http://arxiv.org/abs/1901.02672 (visited on 12 Aug 2019)
4. BBC (2018) Naughty parrot keeps using Alexa to buy things online – CBBC newsround. In: Newsround. https://www.bbc.co.uk/newsround/46566019 (visited on 20 Aug 2019)
5. Chen Y, Cheung ASY (2017) The transparent self under big data profiling: privacy and chinese legislation on the social credit system. SSRN scholarly paper ID 2992537. Social Science Research Network, Rochester. https://papers.ssrn.com/abstract=2992537 (visited on 20 Aug 2019)
6. Chesney R, Citron DK (2018) Deep fakes: a looming challenge for privacy, democracy, and national security. SSRN scholarly paper ID 3213954. Social Science Research Network, Rochester. https://papers.ssrn.com/abstract=3213954 (visited on 02 Sept 2019)
7. Day M, Turner G, Drozdiak N (2019) Amazon workers are listening to what you tell Alexa. In: Bloomberg.com. https://www.bloomberg.com/news/articles/2019-04-10/is-anyone-listening-to-you-on-alexa-a-global-team-reviews-audio (visited on 27 Aug 2019)
8. Defense Science Board (2017) DSB task force on cyber supply chain. https://www.acq.osd.mil/dsb/reports/2010s/DSBCyberSupplyChainExecutiveSummary-Distribution_A.pdf (visited on 31 Aug 2019)
9. E-identity (2019) https://e-estonia.com/solutions/e-identity/ (visited on 02 Sept 2019)
10. Fake Obama Created Using AI Video Tool (2019) https://www.bbc.com/news/av/technology-40598465/fake-obama-created-using-ai-tool-to-make-phoney-speeches (visited on 02 Sept 2019)
11. Fikse TD (2018) Imagining deceptive deepfakes: an ethnographic exploration of fake videos. Master's thesis, p 58
12. Floridi L (2018) Artificial intelligence, deepfakes and a future of ectypes. Philos Technol 31(3):317–321. ISSN:2210-5441. https://doi.org/10.1007/s13347-018-0325-3 (visited on 02 Sept 2019)
13. Henderson C (2019) Package delivery! Cybercriminals at your doorstep. https://securityintelligence.com/posts/package-delivery-cybercriminals-at-your-doorstep/ (visited on 12 Aug 2019)
14. Homoliak I et al (2019) Insight into insiders and IT: a survey of insider threat taxonomies, analysis, modeling, and countermeasures. ACM Comput Surv 52(2):1–40. ISSN:03600300. https://doi.org/10.1145/3303771. http://dl.acm.org/citation.cfm?doid=3320149.3303771 (visited on 12 Aug 2019)
15. ICO (2018) ICO data security trends Q4 2017-18. https://ico.org.uk/media/action-weve-taken/reports/2014675/data-security-trends-pdf.pdf (visited on 29 July 2019)
16. ISACA (2019) State of cyber 2019. https://view.ceros.com/isaca/state-of-cyber-2019 (visited on 29 July 2019)
17. Kelion L (2014) Xbox one ad switches on consoles. In: BBC news. Technology. https://www.bbc.com/news/technology-27827545 (visited on 27 Aug 2019)
18. Lee D (2018) Amazon Alexa heard and sent private chat. In: BBC news. Technology. https://www.bbc.com/news/technology-44248122 (visited on 27 Aug 2019)
19. Lehman J et al (2018) The surprising creativity of digital evolution: a collection of anecdotes from the evolutionary computation and artificial life research communities. In: arXiv: 1803.03453 [cs]. http://arxiv.org/abs/1803.03453 (visited on 21 Aug 2019)
20. Mallmann GL, Gastaud Maçada AC, Oliveira M (2018) The influence of shadow IT usage on knowledge sharing: an exploratory study with IT users. Bus Inf Rev 35(1):17–28. ISSN:0266-3821. https://doi.org/10.1177/0266382118760143 (visited on 01 Aug 2019)
21. Mansfield-Devine S (2016) The imitation game: how business email compromise scams are robbing organisations. Comput Fraud Secur 2016(11):5–10. ISSN:1361-3723. https://doi.org/10.1016/S1361-3723(16)30089-6. http://www.sciencedirect.com/science/article/pii/S1361372316300896 (visited on 12 Aug 2019)

22. Maras M-H, Alexandrou A (2019) Determining authenticity of video evidence in the age of artificial intelligence and in the wake of deepfake videos. Int J Evid Proof 23(3):255–262. ISSN:1365-7127. https://doi.org/10.1177/1365712718807226 (visited on 22 July 2019)
23. Matern F, Riess C, Stamminger M (2019) Exploiting visual artifacts to 17 expose deepfakes and face manipulations. In: 2019 IEEE winter applications of computer vision workshops (WACVW), pp 83–92. https://doi.org/10.1109/WACVW.2019.00020
24. The PEOPLE, Plaintiff and Respondent, v. Terry CHILDS, Defendant and Appellant (2013) Court of Appeal, First District, Division 4, California
25. Ponemon Institute and Accenture (2017) 2017 cost of cybercrime study. https://www.accenture.com/t20170926t072837z_w_/us-en/_acnmedia/pdf-61/accenture-2017-costcybercrimestudy.pdf (visited on 12 Aug 2019)
26. Reed T, Geis J, Dietrich S (2011) SkyNET: a 3G-enabled mobile attack drone and stealth botmaster. In WOOT, pp 28–36
27. Rios B, Butts J (2018) Black Hat USA 2018. https://www.blackhat.com/us-18/briefings/schedule/#understanding-and-exploiting-implanted-medical-devices-11733 (visited on 27 Aug 2019)
28. Sanzgiri A, Dasgupta D (2016) Classification of insider threat detection techniques. In: Proceedings of the 11th annual cyber and information security research conference on – CISRC'16. The 11th annual cyber and information security research conference. ACM Press, Oak Ridge, pp 1–4. ISBN:978-1-4503-3752-6. https://doi.org/10.1145/2897795.2897799. http://dl.acm.org/citation.cfm?doid=2897795.2897799 (visited on 12 Aug 2019)
29. Thing VLL, Wu J (2016) Autonomous vehicle security: a taxonomy of attacks and defences. In: 2016 IEEE international conference on internet of things (iThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CP-SCom) and IEEE smart data (SmartData), pp 164–170. https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2016.52
30. Vamosi R (2018) Casino's aquarium leaks high rollers' personal data, 17 Apr. https://blogs.synopsys.com/from-silicon-to-software/2018/04/17/casinos-aquarium-leaks-high-rollers-personal-data/ (visited on 27 Aug 2019)
31. Wakefield J (2017) Burger king ad sabotaged on Wikipedia. In: BBC news. Technology. https://www.bbc.com/news/technology-39589013 (visited on 27 Aug 2019)
32. Yang X, Li Y, Lyu S (2019) Exposing deep fakes using inconsistent head poses. In: ICASSP 2019 – 2019 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 8261–8265. https://doi.org/10.1109/ICASSP.2019.8683164
33. Zollhöfer M et al (2018) State of the art on monocular 3D face reconstruction, tracking, and applications. Comput Graphics Forum 37(2):523–550. ISSN:1467-8659. https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13382 (visited on 12 Aug 2019)