TAS-S video (Prof. Neeraj Suri)
Transcription and alternative text for the images/slides referred to in the video.

00:02 Hi, my name is Neeraj Suri from Lancaster University and I'll be talking about the security node within the TAS programme.

(**Figure 1:** structure of the TAS network. This shows the a square titled “TAS HUB” surrounded by the 7 nodes which make up the network: Functionality, Verifiability, Legality, Trust, Responsibility, Resilience, and Security which is highlighted by a red circle).

00:10 As you have probably seen from the figure on the bottom left hand side (Figure 1) Security is one of the seven nodes within the TAS Autonomous Systems Programme. And what I will do for the next 20 minutes is give you a simple overview of:

- What are the different issues involved in the security for autonomous systems,
- What is the approach we are following in this node,
- A discussion on what are the different things we are doing within the node.

00:39 So let us get some context and terminology set for context. There are many different notions that people use for autonomous systems. We utilize a very abstract definition that any technology to effectively conduct a mission with varied levels of absence of human intervention is what we define as an autonomous system. So it could have partial human intervention or it could have the extreme of a completely autonomous system. Both of them are within the remit of this node.

(**Figure 2**: The acronym CPS is highlighted in red at the top of the slide, with the following terms listed underneath: Sensors, Communication, Control, Coordination, Navigation, Decision, Adaptation. An arrow can be seen to the right of this list, which groups these terms together as “Perception, Cognition, Decision).

01:13 Now the focus of the application domain for our node is autonomous vehicles. Whether they're terrestrial, aquatic or aerial. And fundamentally, these systems are cyber physical systems. If you notice, CPS is written out here (Figure 2) which is an accumulation of sensory information. Communication aspects control the coordination, navigation, decision, adaptation and many more. But in the community, this is called the PCD approach which stands for perception, cognition and decision to cover all this functionality.

(**Figure 3**: Three summary points are highlighted on the slide. These are: 1). Increasing complexity of applications and environments, 2). Complex connectivity and complex data streams, 3). Cognitively/Computationally complex PCD to AI).

(**Figure 4:** A series of four pictures, laid out horizontally to the left-hand side of the slide. These are (from top to bottom): 1). A single drone in flight, 2). A birds-eye image of five cars connected by wi-fi signals, 3). An image of a car, surrounded by wi-fi signals, stopping at a zebra crossing to allow a human to cross and 4). Several different drones and autonomous air vehicles in flight, connected by wi-fi signals.

01:51 And there are three summary points that I would like to emphasise here, this is especially for our node. (Figure 3.1) An increasingly complexity of applications and environments that we have to deal with. If you look at the picture on the left-hand side, we start with simple, aerial drones (Figure 4.1). Then we deal with a collection of vehicles (Figure 4.2) and if you look at the third element in the bar on the left-hand side, the interchange of vehicles is now interspersed with humans in the loop (Figure 4.3). And the final one is a very ad-hoc environment where entities, swarms of drones or the other, could be interacting with each other (Figure 4.4)

02:32 This leads to the second point in my list at the bottom where it involves complex connectivity and multiple and complex data streams (Figure 3.2). Thus, the net result of this is that cognitively the data streams the collection and the interpretation and the computation are extremely, extremely complex, leading to typically the use of artificial intelligence to solve most of these kind of techniques. It's basically a computational and complexity issue rather than a technology issue.

03:05 So when we come to trustworthy autonomy, let me work around the definition of the word trustworthy. As I have already indicated the complexity is very high in these systems and a simple reality is the more complex the things a system is, things will break, perturbations will happen and this is over all the levels of the cyber physical system. Whether it's the autonomous system, physical hardware by itself, the operations or the environment in it, it works well.

03:38 So the community uses 2 definitions. If the perturbations are coming based on design basis, it could be bugs, it could be misspecification, it could be dealing with lots and lots of edge cases which one cannot anticipate.

03:53 All in the beginning or if the perturbations are operational in nature based on stress that again exceed the design limits or the other. We typically call it dependability. However, if the perturbations are coming through and through bad actors or a deliberate disruptive intent behind it, then we typically use the word security around it. But irrespective of dependability or security, the overall objective is that the autonomous systems acceptably delivers the mission, and obviously the notion of acceptability is going to be defined by the application or the mission. If the autonomous system is operating in a human environment, there are different societal, legal or regulatory issues that will apply. Obviously, if it is a surveillance drone in a different environment, there's a different notion of acceptability.

04:49 Fundamentally, there are two things that we really care about. Whether it's for dependability or security. One is for lightness, that the system does something good, something useful. It's an availability argument. The second one is the more important one, which is called the safety argument that you basically try to make sure that something bad is not going to happen irrespective of the perturbation that has taken place in the system.

05:18 So again, let me do a progressive summary. Complexity is reality in autonomous systems. Uncertainty, especially for the environment in which the autonomous systems work is reality. The use of AI is reality. Where the limits of computation are beyond those of human capability and the typical need in any autonomous systems. And again, I'm not using the emphasis on security and talking about any autonomous system. The basic intent is we want to have predictable behaviour despite all the uncertainties that the system goes through.

(**Figure 5:** A text box with header “The AS need, predictability over/despite uncertainty” and contains the following text: AS depend on technology to provide and improve upon the human capabilities, acceptability and regulations to deliver the PCD functionality. This is a very hard problem)

05:58 And the box (Figure 5) indicates a summary. The autonomous systems depend on technology to provide and improve upon the human capabilities. Acceptability and regulations to deliver the cognition, and decision. And it's no surprise that this is an exceedingly hard problem that the community is trying to solve.

(Figure 6: A text box with header “AS and Disruptions”. and contains the following text: AS expected to for secure [acceptable] PCD delivery with degraded or compromised AS systems [increased uncertainty] This is an orders harder problem.)

06:20 Now by the time we take the same autonomous system and we add disruptions (Figure 6) whether these were designed disruptions, operational or deliberate. The system is again an increasingly complex, increasingly interconnected, increasingly attackable system. And our expectancy now is we still wanted to provide secure and acceptable PC delivery. Except the basic mechanisms, whether it's the computational communication, perception, cognition, has now been degraded based on the amount, based on the perturbation that took place. So fundamentally, we have increased the complexity and uncertainty that the autonomous system has now to deal with to provide the same secure behaviour. So if the previous problem was the basic autonomous system was a hard problem, this is an orders harder problem that security or disruption adds to the whole system.

07:18 Now, each time I use the word security, do keep in mind that I'm talking about the whole autonomous system. That this the security of the entire system that is not compromised. So we aim for the fact that the assets do not get compromised, the operation and the mission does not get compromised, and

the users and the usage environment does not get compromised. And again as before, it's the mission that defines the level and acceptability of what these compromises are going to be.

07:51 So let me start focusing directly on security. So let me give you again in a very abstract view of security in any system, autonomous or computing or software. Security works on the basis that if we are given a given set of assumptions or what are the things that can go wrong in a system. And if we are given a model of the system in which we are supposed to operate, then we are able to assert a very specific security property, whether that's of confidentiality, integrity and availability. And this is the basis that we hope that the real deployment of this system conforms to the assumptions and the model that we had given out there.

08:36: So what are our assumptions behind these assumptions? We have made an assumption that all the assumptions that we made are valid and complete. That the model of the autonomous system, the assets and the environment and the operations are equally valid and complete, and the whole autonomous system and the environment, including the attackers, behave as model. And in under these circumstances the compromise of is what we will call a security breach. Now if you step back, this is a very theoretical, very abstract view of the world, which is perhaps not close to reality. And if you look at the fact that we are making assumptions on not only the autonomous system, the environment, and the behaviour of the attacker, an attacker really doesn't care one bit about what your assumptions and models are. In fact, the more you state the assumptions and the models, the easier the attacker can use, abuse or ignore or do whatever with your assumptions and models. Models are not reality and these are the attacker is not bound by any requirements of either ethics or morality or legislation what it can or cannot do. So models and reality are usually two fundamentally different things.

10:00 So now let's put that in the context of autonomous systems. Do we have accurate and complete models of the system, mission, and environments? I think you know the answer. Do we have complete information on these sensory streams that we are going to be dealing with similarly for the perception, cognition and decision basis? Do we have both accurate and complete information? And so on. And basically the autonomous system is a world of uncertainties where we are working in a very complex, very dynamic and very uncertain environment. And this is where we want to provide some very predictable properties.

10:40 So I'm jumping ahead in my slides out here to again make the assertion that security is something that works very well in structured environments when we know the model. When we know the assumptions, and autonomous systems are anything but structured. So the basic problem becomes how do we go about providing well structured, very predictable behaviour of autonomous system security when we are dealing with a very complex world of humans and objects or the others with the mission environment and

everything is fundamentally unstructured. So an extremely difficult problem and again it would be silly to say that we don't know how to solve it. Of course we know how to solve these problems, but we know how to provide very specific and very expensive point solutions to very specific missions, whether it's for cars or drones. Whether it's a commercial mission or a military mission, these are very specific point solutions.

11:43 So what we want to do in the TAS project is basically go beyond point solutions to develop a scientific framework which can provide a composable, scalable and verifiable mission adaptive approach to security. So again, it's predictability, despite uncertainty, which is the driver behind the entire node.

(**Figure 7**: A chart showing the three research strands (RS) of the TAS-S Node. These are (from left to right) RS1 “Can we secure the AS usage basis?”, R2 “Can we secure the AS operations?” and RS3 “Can we secure the RS users?”)

12:09: Now this is a very multidisciplinary, or rather an interdisciplinary node. I have only given you the background that we are dealing with assets, operations and users (Figure 7). Now what the node does it very specifically addresses all of these three issues as first class citizens. So again, let's take the premise we are dealing with unstructured, uncontrolled and dynamic environments. And our intent out here is threefold from a basic usage viewpoint. We have the first trend where we look at the fact “can we secure the usage basis of an autonomous system”? Then we look at the second strand the middle, which we call “can we secure the operations of the autonomous systems?” Whether it's for the mission, for the operations, for the control, for the navigation, the information and communication. And the third strand is again very intertwined with the first two technologies strands where we specifically focus on the users by saying “can we secure the users of the autonomous system environments? And these are the ones which can get affected. So losing a drone or losing a car is obviously not a good proposition. But if the loss or any damage to these assets end up hurting the people in the environment or causing collateral damage, obviously that is not an acceptable thing. So we consider the usage, the operations and the users on a standalone basis within the node.

(**Figure 8:** This is a slide titled “Addressing “uncertainty” and Functionality”. The following bullet points are listed underneath: 1). Specification of the AS systems environment, 2). Specification of discrete/computational AS security (SOS), 3). Form, adapt and solve AS’s PCD models on the fly (FedML), 4). Develop explainability and verifiability od AS’s PCD behaviour. AI is wonderful but deterministic reproducibility is not its strength.

13:44 So let me elaborate a little bit on each one of these. So for the first trend where we focus on the autonomous system environment (Figure 8.1). One of the key problems that we saw was is it is extremely hard to specify what the entire systems environment of the autonomous system is. So starting with that kind of specification, we come to the next harder problem which is the

2nd bullet (Figure 8.2), the specification of autonomous system security. Now I've written the word SOS in brackets, which stands for system of systems. So whether it's going to be a single entity or multiple entity, we have to deal with discrete security and the compositional security. So if I say I'm flying, a military mission is flying a certain number of drones and they have to interact with many other vehicles on the ground, in the air, or the other, it's a system of systems we are trying deal with. Now, specification only leads to the basis for the PCD, where we basically define different types of federated machine learning. Architectures and algorithms on how one would go about solving the perception, cognition, and decision problem in these complex environments. And as I mentioned earlier in the talk, we will be using AI, that is a natural use in the autonomous systems area. Now what AI is wonderful but determinism and reproducibility is not something that AI is particularly strong at. So when we develop the solutions with the use of AI, the explainability and verifiability of the AI autonomous system behaviour is something that we treat as a first class attribute in the project.

(**Figure 9:** This is a slide titled "the PCD Ops: Threat modelling and mitigation". The following bullet points are listed underneath. 1). Ascertain attack surfaces, i.e. entry points and likelihoods across the PCD mission plane, 2). Provide quantifiable feedback to mission space ensure that limits of controllability are not compromised 3), Provide secure communication across/to the AS' information plane.

15:32 The second strand, where we deal with the PCD operations (Figure 9) is something that focuses on the threat models and mitigation of the security attacks that place in the system. So we are obviously looking at ascertaining the attack surfaces over the entire mission plane. We are looking at the control and the navigation plane to make sure that the mission space and the limits of controllability of the autonomous system do not get exceeded. And the heart of the whole of an autonomous system interaction is communication, both for basic communication and also for exchange of the data streams that go behind it. So specific techniques to provide secure communication across the autonomous system and to the autonomous systems information plane is what this trend emphasizes on.

(**Figure 10:** This is a slide titled "RS3 The Users/Societal Space. What roles do human, ethical, legal and environmental factors play in AS security?" The following bullet points are listed underneath: 1). How do human behaviours and social factors drive AS security? 2), What ethical, regulatory and legal factors need to be considered over secure AS ops? 3). Can human and technical factors be integrated to result in preventative secure AS design?)

16:25: Last but not the least, as I indicated earlier, autonomous systems operate in societal spaces, so the major interest out here is what roles do human, ethical, legal and environmental factors play in autonomous system security? I'm not going to go into the details, but I'll give you a very simple example. If a series of let's say, autonomous vehicles, have to make a decision that if, based on a security attack the vehicle is going to bear to the side. It could

damage other vehicles or it could damage a building or it could damage or it could hurt people. The people could be a group of elderly people or a group of children. How does the system on the fly in the presence of an attack make a decision? On what collateral damage, whether it's a it's a physical damage, a human damage is acceptable? And these are socio-ethical problems that one has to deal with.

17:25 Now, one aspect we also look at, the fact is that can human behaviour be adapted or be sensitized to the use of technology in these kind of scenarios? So we look at the fact how do human behaviours and social factors drive security? Overall, whatever basic interest is, can human and technical factors both be integrated in the design and operation of the autonomous systems? Now obviously we are dealing, we are starting from the foundations but the whole project is based, it goes from theory to practice where we have a whole series of actual test beds both at Lancaster and the partner Cranfield University where we deal with drones, we deal with threat attacks, we deal with communication test beds. So there is a whole range of test beds and datasets that we are developing or have access to in the project and this might be a good time to say that we plan to make all our datasets public to the rest of the nodes and the Community as possible.

18:30 In terms of outreach and engagements, I would very strongly emphasize to look at our website . We have a lot of activities in terms of workshops that we have conducted with our multiple stakeholders. There's a whole series of events we have held in the past and they will be upcoming in the future. All of our seminars, internal and external or public. Internal are the ones which are hosted at Lancaster or Cranfield and a number of industrial partners are also presenting their talks. Please go to our website, these are all public and open to all. We have a lot of exposure to industry, government and regulatory bodies. We will be conducting of a number of services and we have a and we have a very active dissemination, photos with blogs, Twitter and newsletter. Again, the emphasis is to go to the website and you will find many opportunities and mechanisms to engage with us.

(**Figure 11:** This is a slide showing the names and photos of the TAS-S Team, based at Lancaster and Cranfield). This information can also be seen on the TAS-S Team webpage.)

19:37: Last but not the least, our team is very interdisciplinary in nature. Spread across Lancaster University and Cranfield University, we have a very interdisciplinary profile: system security, machine learning network, social sciences, law, behavioural sciences. And again, the heart of the project is our students and postdoctoral assistants, and all this information is on the website. And driving the whole thing is a whole bunch of stakeholders. We have started from Airbus, BAE, Raytheon. (A full list is available on the stakeholders page on the TAS-S website). There are multiple universities in England in Europe and across the world. And we have a very comprehensive base of stakeholders covering both civilian and military applications. Our interest is basic scientific foundations that we're developing and we are very fortunate to have an Advisory Board which spans the whole globe and we

have the hub liaison team from Kings College, Nottingham and again, this from King's College that play a very active part of the whole project.

Thank you.