Digital Identity – The Ghost in the Machine

Max Kilger
Chief Social Scientist
Honeynet Project

[Show Prisoner Video]

Introduction

One of the more critical social psychological elements associated with personality, motivations and behaviors is the concept of identity – e.g. having a sense of one's own identity. Indeed a sense of one's identity is also an important component necessary for the psychological well-being of an individual. Therefore I would like to suggest that it is with some serious thought that we examine how digital technology is having a significant effect on both how others see us and how we see ourselves and the resultant effects on identity. In this paper we will treat the concept of identity as a multi-dimensional social element and examine how digital technologies such as the computers and the Internet affect personal identity.

Identity as Process

The formation of personal identity is both an ascribed as well as a constructed process. When you are born there are some simple aspects of your identity that are typically assigned to you very quickly. In the most simplistic terms you are assigned a gender – male or female – and a name – Helmut Kilger, for example. While these processes are simplistic in nature, their importance can be underscored if there is somehow a disruption of this process.

For example, after the birth of a child there is a short period of time during which it is normative the child may not have a given name. However after that period of time has expired, if the child still does not have a given name, friends and family will begin to inquire more ardently about the given name of the child. Indeed, wait long enough and the government or state will insert itself into the identity process and demand a given name be assigned to the child.

Gender assignment is also an element of identity ascribed at birth. Again we have inquisitive friends and family inquiring about the gender of the child – any disruption of this identity assignment is going to be met with social pressure. Eventually again the state will intervene as well and demand to know the gender of the child – and if there appears to be some doubt about it, the new parents are often forced by social and legal pressure to choose a gender.

[Do the switched at birth routine here]

From the point of birth onward, your identity is formed through a very large and continuing number of processes. These processes may be rooted in social, legal, commercial or other realms. Let's take one of these processes that shape social identity and see how technology might have an effect on the outcome.

Status processes occur all around us – they everyday and involve everyone in this room. Status processes involve comparing information about characteristics we possess with the characteristics that other people possess. These status characteristics might include demographic variables such as gender, age, education, income as well as variables related to performance – such as Heidi there is a much better C++ coder than I am.[1]

How do these status characteristics get noticed and observed? Often this comes in face to face interaction – you sit down with Heidi to talk and notice she is female, learn she's 5 years older than you, has a Ph.D. while you have a Hochshule diploma and she shows you some C++ code she has written that is much better than anything you have ever written. You process that information in a mathematical manner that I won't go into now and adjust your identity in relation to Heidi. In a professional dimension, perhaps you begin to think that you are not quite as good a programmer as you thought you were. Also maybe you begin to re-evaluate yourself in terms of attractiveness as a mate as well.

This is a much more difficult process when done online. First of all, you can't directly observe that Heidi is a female – you can only assume that she is. You have some clues to go on – her name is Heidi – and perhaps you even have a picture of her online. But how do you know that Heidi is her real name? It might really be Johann – and the picture of Heidi might not be the person you connected to on the net as well. Also, how do you know that Heidi even wrote that piece of code she sent you?

So how do you gather evidence that determines Heidi is who she says she is? In real life face to face interaction these status characteristic clues we've discussed are collected and then matched against verbal and non-verbal clues that reinforce the information you've already gathered. In face to face situations, people are voracious information gathers. People are constantly collecting information about others, evaluating it and incorporating it into the identity of the person they are talking to as well as adjusting their own identities in relation to others. Let me demonstrate what I mean.

[face to face interaction demo with audience member]

These kinds of verbal and non-verbal cues are generally not available in typical communications that occur over digital networks like the Internet. Now some may argue that this problem can be alleviated by using video conferencing, webcams, etc. However, we know from research that even video conferencing degrades the ability of people to read, interpret and evaluate these verbal and non-verbal cues. There is a reason, after all why I flew 10 hours here to be in person rather than just appear on a big video screen in front of you.

Identity as a Temporally Unstable Element

One of the key dimensions of identity is persistence. Every element of a person's identity has its own independent timeline. Some elements such as occupation can have slow timelines –for example, you may work at the same occupation for many years or you may have the same national identifier such as a social security number for the remainder of your life. Other elements of identity have quicker timelines – something as fleeting as the 60 second life of the password generated by this RSA token [hold up token]. Indeed, the very idea that the formation of identity as a process strongly suggests that the identity is temporally unstable.

Online identities are a special instance of temporal instability. Computers and networks make it possible to create an almost infinite number of commerce, security and communal systems and have individuals create or have assigned identities in each of them. These identities are often fleeting – after a period of time the person retires from the community to which they have created the identity, or they are forcibly removed or their identity expires and this part of their digital identity ceases to exist. This suggests that people often maintain numerous identities and that these identities are always in flux from a temporal standpoint. At any point in time there is some probability p that the identity will cease to exist in time period t and a new identity emerge with probability q in time period t+1.

Identity persistence is also important in the information security arena. Identity persistence means that the elements and details of identifying, authenticating and authorizing actors require reliability – the characteristic of repeatable results. That is, the criteria must either remain stable for some period of time t or must change only in a predictable way within that time period (as is the case with the RSA token).

The temporally unstable nature of identity has at least several potential consequences. First, an individual must be able to adapt to possessing multiple identities which change in perhaps not completely predictable ways over time in order to remain a healthy, well-adjusted person. Secondly, it poses potential challenges to identity management architectures (IMAs) in that one may not always count upon identity

persistence over time - a key feature of many IMA schemas - something that suggests that there may be "hidden cracks" in the developing area of inter- enterprise identity management systems.

Situational Identity

Social scientists who study identity have long observed that identity in a social sense depends upon the situation in which the actor finds him- or herself. For example, a person may find themselves assuming the identity role of father to his son at home while assuming the role of colleague in his work environment. This suggests that individuals can adopt and manage numerous identities and activate them when they find themselves in the appropriate situation.

Similar phenomenon can be found in the digital world. The identity that an individual takes on depends upon the situation that individual finds themselves in. For example, someone might take on the role of a orc-battling hero in an online game and then later that same hour sign on to Travelocity.com and assume the identity of a traveler to Berlin. The idea that identity is situational in both the offline and online worlds is not a new one but there are lessons that can be learned here.

In the example just cited, both identities are compatible because they exist in distinct and discrete situations and these situations do not communicate with each other. I bring this point up because there is a very good paper on identity that people are probably quite familiar with dealing with the concept of an augmented social network (ANS).[2] The authors of this hypothetical identity system argue that there are four elements to the system: 1) persistence of identity 2) interoperability between online communities 3) brokered relationships and 4) public interest matching technologies. We have already seen that there are some potential cracks in the persistence of identity element. In addition however, there are also some cracks in this scheme due to issues of situational identity.

The ANS system relies to a non- trivial extent upon trust relationships in several senses. In the first sense it relies upon trust when one actor brokers a relationship between two other actors. More to the point, the ANS system relies upon trust identity across online communities. That is, a trustworthy actor in one community is a trustworthy actor in another online community. While there is some truth in that in some sense there a "halo" effect that an actor that you trust in one community can be trusted in a second community, this idea of generalized trustworthiness can easily get you in trouble.

As a counterexample, I personally know someone who is an excellent researcher and in the field of research methodology I trust her judgment implicitly. Jordan, Hauser and Foster suggest in an ANS schema you

could transfer this trustworthy reputation or trust relationship to another online community – I'll pick Ebay for example. Would I trust this person as a seller or buyer on Ebay? Most definitely not, because while I trust her in the research community I certainly would not in the Ebay community because although there is that halo effect I know from personal knowledge that my friend is simply terrible at paying bills, has her cable cutoff for nonpayment, forgets to make insurance payments and all other forms of fiscal mayhem. And it's not because she is untrustworthy in general – it's because in this particular community – Ebay – I know I'm never going to see my money from her – he attention is elsewhere.

Therefore we must be careful about examining identity because to a non-trivial extent identity is situational and situations can significantly alter the likely outcome of events. This also suggests that situational identity is an issue that must be dealt with in emerging identity schemes such as federated identity systems.[3]

Identity – Class versus Unique Identifiers

In order to identify someone or for someone to identity themselves or for someone to evaluate their own identity, some sort of hopefully universally understood identifiers or claims need to be associated with the individual. There are several attributes within this dimension of identity that should be at least briefly mentioned and discussed.

The first attribute involves the distinction between a class versus a unique identifier. For example, you may identify yourself or be identified as having been issued national id number 53307507484. To the extent logistically possible, this is a unique identifier that points to a single individual and the assumption is that no other individual can be associated with this identifier.

Your identity may also be linked to a class of identifiers for example, you might be a college graduate, female, age 34, a bunny rabbit owner. These are class attributes that you share with at least one other individual and so do not uniquely identify you. For some purposes such as going to the restroom, knowing your identity as female and having others accept that identity is a pretty useful thing.

In the online world we have the opportunity to fairly easily adopt identifiers. It's possible – and I am sure there a number of folks in the audience who have done this – to pose as someone of the opposite sex. You could also adopt the identity of someone who works for a government agency and probably get away with it, at least for awhile.

The key difference I want to make here is in comparing the traditional physical versus digital world. In the physical world it is possible to adopt

a new identity as somewhat younger or older with a bit of makeup, become female with more makeup and some costume changes, adopt an identity of being college educated with a bit of study, etc. However, it is significantly more difficult to switch identities for things that are unique identifiers such as national ID cards or using a credit card belonging to an individual, although this is done.

In the online world this gap is much smaller. Because there are fewer or no physical clues, because there are always new technologies coming out that seem to help solve technical issues and so the gap of difficulty between being able to assume a class identity and a unique identity is much smaller in the online world.

You can even use digital technology to manufacture props that can be used in attempting to adopt new identities in the offline world. You can create websites that contain information about your identity and eventually search engine crawlers will find that information, link it to you (usually as a unique identifier such as a name or nickname) and index it. People have a tendency to legitimize results that they receive from searches using well-known search engines such as Google and so the props you planted help become part of your identity.

The point here is that in the digital world the small gap between what is real in a class sense and what is real in a unique identifier sense loosens our hold on self identity and who we are – in the digital world we can really alter class and unique identifiers so efficiently that we may have lost a bit of our own sense of who we are. In essence we have brokered a world where one can manufacture identities of convenience and there may be a psychological price to pay for that bargain. As we create and have created for us more and more identities I believe this may denigrate our ability to synthesize these alter ego identities into a single sense of who we are – and this may not be such a good thing.

The second attribute that I want to discuss has to do with the comparison of deterministic versus probabilistic identifiers. Deterministic identifiers are those characteristics in which we feel have a high level of confidence in identifying ourselves. For example, you can with great certainty know if your identity includes being a rabbit owner – you have close, personal knowledge of whether or not this is the case. Similar arguments can be made for other simple demographic variables such as age, gender, education, etc.

There are certain aspects of your self-identity that have a probabilistic rather than deterministic one. These might include, for example, identifying yourself as a good driver or bad dancer. The probabilistic portion of this is that you do not know for sure this is the case and your

ability to reference yourself in relation to others is often limited. Often these probabilistic self-identified attributes involve modifiers such as "good" or "bad" because it is the very nature of the uncertainly in self-assigning these adjectives that makes them non-deterministic.

The distinction of deterministic versus probabilistic takes a much more interesting turn when you put digital technologies into the mix. For example, imagine that part of your identity that you have associated with yourself is that you are an expert C++ programmer. You are even likely to communicate that over the net to friends in a virtual social group or personal website. However, now complications arise because your reference group is now much, much bigger – in fact, it is the rest of the online world. It is now much more problematic to own the identity of expert C++ programmer because expert is relative and now your comparison base has exponentially grown from a small group of people you know and have physically met to literally the entire online world..

In addition, because of the lack or limited nature of verbal and non-verbal cues in most web-based communication it is difficult for others to evaluate your identity as an expert C++ programmer. In fact, what often happens is that this lack of cues often results in status conflicts among members of the online community to determine their place in the status hierarchy. These status conflicts are often followed by social control processes like flaming, hijacking machines already compromised by the offending party, etc. One of the ways in which an individual can establish their identity is to demonstrate high levels of competence by writing elegant hacks and code that is evaluated as such by other individuals. Another method in securing one's identity is to physically meet with other people and exchange verbal and non-verbal cues that help establish and solidify self-identity – this is partially what is happening here in Berlin. Hacker conventions serve a very functional purpose in this extent in that they facilitate the exchange of verbal and non-verbal cues with others in a very efficient manner and allow actors to sort out their relationships.

One final discussion about probabilistic identity involves the use of statistical algorithms and statistical models to assign characteristics of identity. In the business world everyone is quite aware that there are large commercial enterprises that collect and integrate large amounts of data about individuals for the purposes of marketing, client service, etc.. While these databases often contain information on hundreds of millions of individual and often have thousands of variables, the data fields themselves are often not well populated. That is, there are many variables or pieces of information about individuals that may be useful to these commercial entities but often they can capture data for these variables only for a small percentage of the individuals database – the rest of the individuals hold missing values for these variables.

One of the ways in which companies resolve the missing data issues is to use statistical algorithms and models to attach propensities or probabilities for individuals for these variables. For example, a marketing company might run a model for all individuals in its databases to predict which ones are likely to watch the CNN television network and which ones are not likely to watch. Thus the identity a company develops for you as a customer or prospective customer may often incorporate characteristics that are based on a probabilistic estimate rather than deterministic information.

Other applications of this type of technology are the databases developed and maintained by government agencies (intelligence, law enforcement, etc.) They use these databases for national security purposes and sometimes it might be necessary to impute or model some variable for a person of interest in order to better understand the risk the person may pose to a specific entity. The idea here is that digital identities are very real and have very real-world consequences for individuals and that you may have an identity attached to you that is based at least partially on synthetic data. It is likely that this trend is only going to accelerate.

Source of Identity Information – Self or Other

It is clear from the previous discussion that not only does an individual form their identity from their own information and evaluation process but there are others out there in the digital world creating identities often without the detailed knowledge of the individual. Do these other-created identities have an effect on the formation of our own self-identity? If you look closely at how George Herbert Mead describes the emergence and organization of the self, he uses two terms – the "**I**" and the "**me**". Mead refers to the **me** as the social self – that component of self and identity that arises from the way in which others interact with you as an individual.[4] This suggests that the identities that others create for you do in fact not only affect you in the physical world but also have a non-trivial effect on your own identity. That is, you yourself incorporate other people's impressions of you into your identity.

Mead describes the **I** as the novel or individualistic way in which the individual reacts to the social self or me. Therefore it is probably safe to say that individuals form at least part of their identity from their individualistic reactions to the identities formed for them by others.

Another way in which digital technology is shaping our self-identity involves our social interaction with computers and computer technology. It is becoming clearer in my opinion that people treat their computer as a social actor – that is, along the theoretical lines of George Herbert Mead they exchange meaningful social symbols with their computers and thus

treat them as social actors. This means that the identity individuals create for themselves in some fashion shaped by the very digital technology they use, not just that digital technology is often a communications channel for content but that the computer itself is considered as contributing social content to the lives of people. In turn, the changes digital technologies bring about in your identity also affect how you interact with other people. These are interesting thoughts and it is difficult and too early to completely understand, in my opinion, what the full consequences of this may ultimately be for us as human beings.

The Future of the Digital Individual

We now know we have literally tens if not hundreds of digital identities out there in the world for each one of us – some of them constructed by us and some of them constructed by others. As more and more of our sense of self beings to come from these digital identities it suggests that we as human beings are becoming *the ghosts in the machine*.

In the future our work and social lives, our intimate relationships, our perspective of the world, our complete identities may emanate from the digital realm. There is already the sense in a small way that meeting people face to face is a bit odd, especially for people like yourselves whose lives are deeply embedded in technology. Meeting someone face to face – that is, really meeting them for a purpose - may someday be a very rare, unfamiliar and awkward event. We may begin to lose the ability to effectively communicate in a face to face world by losing the ability to interpret the verbal and non- verbal cues.

Moreover, the growth in the number of digital identities associated with us as individuals may lead ultimately to the fragmentation of the self – the inability to formulate and retain an integrated sense of ourselves. This fragmentation of our identity into so many different pieces is obviously going to have consequences both for our psychological well- being, and it is going to be interesting to see just how it affects our quality of life.

There are also likely lessons to be learned here about identity and its relationship to information security and digital identity management. It is clear that we are still in the very early stages of trying to develop identity management systems for security purposes. It is also clear that these identity management systems while complex from a technological standpoint are still quite primitive when compared to the complexities of how humans construct and manage their identities. Learning more about how people construct and manage identities may provide some valuable insights in the information security arena. This suggests that concepts like the “Seven Laws of Identity” [5] are an important advancement in digital identity management, there may indeed be a long way to go before we can apply some of the lessons to be learned.

For those of you who fear that the development and construction of digital identities by those other than yourself will lead to negative consequences both personally and for society, I would suggest that there is indeed hope in the fact that we can construct larger numbers and more complex digital identities and shed them faster than others can mine them for information when we construct alter ego digital identities of ourselves. Therefore as individuals we have a head-start in front of the giant jaws that represent organized efforts to collect, analyze and reconstruct our digital identities. However, this does not mean that we should be complacent.

Summary

I hope that you have found, in this paper, an idea or two that stimulates your thinking about the dimensions of digital identity, the sense of self and the future of identity architecture. I'm Dr. Max Kilger – at least I **think** I'm Dr. Max Kilger – and thanks for listening.

1. For a more formal introduction to the theory and mathematics of status evaluation see Berger, Joseph, M. Hamit Fisek, Robert Z. Norman, and Morris Zelditch. 1977. *Status Characteristics and Social Interaction: An Expectation-States Approach*. NY: Elsevier.

2. The Augmented Social Network: Building identity and trust into the next-generation Internet by Ken Jordan, Jan Hauser, and Steven Foster, First Monday, volume 8, number 8 (August 2003).

3. Federated Identity Systems, Whitepaper, Eric Norlin and Andre Durand, PingID Network, Inc.

4. George Herbert Mead, *Mind, Self, and Society*. Chicago: University of Chicago Press, 1934

5. Kim Cameron, Laws of Identity, Microsoft Corporation, 2005.