Everyday Electronic Communication for the 21st Century

by Nigel Hathaway

(version 1.0 dated 19th August, 2017)

Abstract

This white paper addresses the problems that plague modern telecommunications, and in particular the Internet, namely unwanted communications, fraud and identity theft. The cause is discussed, and a practical solution is proposed, along with the issues associated with establishing such a solution.

About the Author

Nigel Hathaway is a freelance electronics and software consultant, currently contracted to a leading telecommunications small cell software vendor. He specialises in Linux driver and system level software, bootloaders, system security and encryption. He has over 30 years experience in the electronics industry, the most recent 10 years of which have been in small cells.

The Problem

The Internet has created a revolution in communication between individuals, groups and organisations in a way that was never envisaged before the late 20th century, or even when the Internet was in its infancy. One of the key factors in encouraging the extent to which this has occurred, is the zero incremental cost of using the available services. Unfortunately this, coupled with the original idea that all users would behave in a decent and honourable fashion, has led to an unprecedented level of attempted fraud and unwanted communication perpetrated via the Internet.

At the heart of the problem lies the issue of the establishment of identity, the maintenance of data confidentiality and weaknesses in maintaining permissions as to who can contact whom. The primary conceptual weakness in current established systems is the use of “shared secrets” such as telephone numbers, e-mail addresses, user names and passwords. Each of these suffer from those secrets falling into the hands of malicious entities, and this is the main means by which fraud and identity theft is facilitated.

Much has been done to tackle these problems, but the solutions tend to be proprietary to particular large organisations, and even then those solutions allow unsolicited contact attempts (e.g. from pornography providers) and many still rely on passwords, which is a notoriously weak and problematic form of securing identity. To use these proprietary systems, every person has to be a ‘member’ and communication is often monitored for advertising purposes.

In summary: how do we eliminate spam, nuisance/fake calls and identity theft?

The Solution

The Internet relies on vendor-neutral standards (embodied in RFCs) to allow universally compatible communication, and it is highly desirable for the solution to follow this rule, so that the whole world does not become reliant on one vendor who can then hold it to ransom.

The following set of features provides the basic foundational principles:

- There should be no shared secrets (such as telephone numbers and e-mail addresses) used to solicit communication.
- The system should not rely on information, known (supposedly) only to the particular individual, for establishment and recovery of identity. Such things include mother's maiden name, favourite colour, childhood best friend's name and so on.
- There should be no reliance on several large, or even a single global, register of identities, due to the consequences of such a database of being compromised. Identity information should be distributed, so that the damage from any security breach is limited.
- Access to the system should not be governed by individuals' ability to remember a large number of awkward secrets which may have to frequently change, in particular passwords. The system should use a smaller number of other features, such as biometrics, current physical location and items that individuals possess.
- The network location of a device on which an individual may be contacted must be readily discoverable by people whom that individual has authorised, and not by others. This must be done without relying on a small number of providers (or even a single one). This should be distributed in much the same way as DNS.

From a technological perspective, the goal is not to invent anything particularly new, but to use existing standards and methodologies as far as is possible.

From a user's perspective, the system will have the following basic features, which are largely familiar from existing systems:

- Each 'identity' has a contact list, where each contact has associated permissions as how they can make contact on each device, e.g. text-only, or text + audio, or text + audio + video (text includes data as per current e-mail). No-one outside the contact list can make contact with that identity. Contacts also have 'metadata' so that they can be sorted and categorised/grouped.
- An identity may represent an individual, or a group, or a piece of equipment such as a fixed line telephone in a home or even the controls of a heating system. An individual may have multiple identities they can operate, and identities may be shared among a number of individuals.
- An identity, with its contact list, may be present on a number of different devices with different capabilities, from simple telephones, to tablets and computers. Each device synchronises with every other with regard to the information it contains for that identity, including contacts, messages and call logs (in so far as that device supports that functionality). Multiple identities may be held on a particular device. It should be noted that, for a device to be able to (securely) hold an identity, it must itself have a secure means of identification, so as to prevent the existence of cloned devices holding stolen identities.

One of the key differences between this and existing systems is the way contacts are added to a contact list. There is no central or searchable directory of identities, and it is not possible to make an unsolicited contact request. However, there are a number of ways contact connections are established, examples of which are as follows:

- Two devices in the same physical locality communicating directly with each other, for example two smartphones in close proximity exchanging details over NFC or bluetooth. An authorisation page is opened on both phones, and mutual agreement is made by the two users.
- A mutual contact of the two unconnected parties makes a contact recommendation. The request appears on the devices of the both the unconnected parties, and when they both agree, the connection is made.
- Contact is made via a contact-making website, where a time-limited one-time-passcode (OTP) is generated and that code is entered into both devices. Clearly, that code needs to be shared by other means.

The issue of loss (or theft) of a device, leading to possible use of an identity by an unauthorised person, is handled in the following way. A device can be registered as compromised by two or more other devices either containing the same identity, or the identity of a specially-trusted contact (which may include organisations such as an ISP). All identities on that device become disabled as soon as one of the reporting devices is able to make contact with it. Clearly, there is the issue of what happens if all the devices used by a particular identity become compromised and disabled (for example, if the user has only one device, which is likely to be quite common). If the information is “backed up in the cloud” then this would typically fall back into the realms of access-by-password which this system is specially avoiding. USB security tokens are an available technology that can assist in this area, but are still comparatively expensive. A much cheaper option is the contactless smartcard (or even chip card), but this has to be used in conjunction with special equipment, which leads on to details of the “ideal” implementation.

The system uses public key cryptography to implement identity. Each ‘identity’ has its own private/public key pair and each device has a private/public key pair per identity stored there. The identity issues signed certificates for the devices that use it. The identity itself has a certificate signed by the trusted system on which its private key is stored, and that secured trusted system has a certificate signed by a root authority. The problems with storing that private key “in the cloud” are (a) passwords (as mentioned above) and (b) mass theft of identities (private keys) from one place. An alternative solution is for the user to have a secure device that is kept “at home” or some other secure non-mobile location. Such a device is able to manage multiple identities, and needs to be network-connected for the management and issuing of certificates. Such a device may, for example, be built into an ISP-supplied broadband router and have a smartcard antenna. The root of the identity itself (a public/private key pair) may be kept on a pair of smartcards (master and backup), minimising the possibility of identity loss.

As well as the storing and maintenance of identity, the communication of one device to another has to be facilitated. If contact is to be made with a mobile device whose IP address is constantly changing, then that device needs to maintain a registration with a system whose IP address does not change. This means that each device/identity combination has to have an Internet address at some sort of “provider”. This service could be provided by the same device on which identity is managed. However, even if it is located at the end of a DSL line which is reliable, there are issues to do with non-static IP addresses (even with IPv6) and the overall complexity of managing firewalls and DNS by the user. Therefore, the “ideal” arrangement is where the ISP also provides the location service, which is a service that doesn’t require much bandwidth or computing resource (similar to ICE servers

To leave the description of the system here would not do it justice. Current telephony systems, including mobile phone systems, have a somewhat unimaginative feature set, which hasn't advance much from when the telephone was invented: in particular, the presence of nothing but a numeric keypad is a limiting characteristic which is not appropriate to 21st century communication.

The protocols used all for extendibility is ways not yet imagined. However, the following features are part of the basic system, and assume at least a smartphone-style screen interface:

- Call hand-over. As well as being able to go to the contact list, and pass the call on in classic PBX fashion, a voice call on a telephone can be transferred to the user's laptop to become a video call without dropping the call.
- Group call / conferencing. This concept is well-established on systems like Skype, and such systems also allow for screen sharing and so on – although people usually only want to share one window for confidentiality reasons. They may also want to allow other parties on the call to take control of the window/program being presented.
- Pre-call data selection. When one person calls another it was once considered an innovation to see the number of the calling party before accepting the call. This system goes beyond that in that a form can be presented to the caller prior to initiating the call, where they can, for example, say what the call is about or how urgent it is. If the call is interrupting the called party, they can then decide whether to accept the call on the basis of that information. Such a facility also eliminates the need for IVR systems, where you press numbers and then eventually end up listening to music before the call is answered. All this information can be entered before making the call, and if there is a queue, you can be alerted to pick up the phone rather than listening on hold. This kind on facility is potentially very useful to the emergency services as well.
- Membership of Organisations (groups). There is a facility whereby, if an individual becomes a member of an organisation, then that membership can automatically allow all members to be able contact all other members of that group.
- Interaction with commercial entities and other bodies. Such organisations need to be able to publish public contact details, for example on their website. When the user (using a conventional computer) clicks on the link to make contact on a website, they are given the option of making the call from their phone instead of their computer. In addition, the user can add that organisation to their contacts if they want to.
- Mailing lists. A mailing list can be an 'identity' in a contact list. The user can then control what they receive from that list, and the list can also control what is posted to it.

Strengths, Weaknesses, Opportunities and Threats

Any communication system has the chicken-and-egg problem of needing multiple subscribers before it can be used. For people to subscribe, there needs to be perceived benefits, or they won't make the effort. Because all similar systems are essentially free of charge, the entry level of this system also needs to be free. Thus, the basic subscription needs to avoid having equipment that needs to be purchased. Ideally, if equipment is involved (namely the device holding identities) this needs to be provided free-of-change to the end user by organisations who have a vested interest in their having it (banks for example).

Although this system is aimed at obsoleting the subscriber number-based telephone system, that system will remain for quite some time. Currently pricing is quite competitive between conventional telecommunications companies and VoIP providers. However, VoIP is very under-used due to the equipment and complexities of setting it up. This may be an instance of where it could be in the interest of other parties to promote they system. For example, one possible approach is to encourage mobile subscribers to have (cheaper) data-only subscriptions, and to make all calls via lower rates, particularly to landlines – with calls to their contacts being free, of course.

From a business model perspective, the protocols and system are open (and open source), this will encourage third parties to maintain their own client software. The primary business opportunity lies in certification, and associated secure equipment. For proper protection against identity theft, the private keys associated with identities need to be held on secure hardware. That hardware must have secure boot (i.e. only boot signed software) and must ensure that private keys cannot be leaked.

Commercial entities and other bodies need to ensure that they cannot be impersonated by fraudsters on the system, therefore there is a need for “levels of certification”. For basic friend-to-friend contact, this is not necessary. But banks, for example, would want to ensure that no-one giving themselves the same name (or close equivalent) ended up on someone’s contact list. So, the certification level of each contact needs to be clearly marked. Similarly, individuals may need to have their identity certified in order to be able to deal with banks and other organisations. Certifiable information includes the person’s full name, their date of birth, their address and so on, and the individual would need to be able to choose which certified information they shared with whom.

The “Internet of Things” is currently a trendy top, and a classic example of this is a central heating control thermostat, controlled remotely. These already have a reputation for security breaches and outages, and this is because the individual devices rely on external Internet severs (which may go offline permanently anyway during the device’s lifetime). As part of the system described here, IoT devices need not rely on external servers, and will therefore not have the vulnerabilities described.

Finally, one thing that should be mentioned in the concept of “lawful intercept” because, not taking it into proper consideration could lead to the system being banned in certain places. Therefore, it is better to build provision for it from the start, even if it is initially disabled. Most countries require “telecommunications providers” to provide a back door to allow them to monitor communications under certain circumstances. However, this system is peer-to-peer, and there is no “provider” as such, so it effectively bypasses most counties’ current laws in this respect. End-to-end encryption is needed because of the existence of open WiFi (which would otherwise allow anyone to listen in). This is a thorny issue with no obvious answers, and which has the potential to reintroduce fraud.