The Internet Telephony Red Herring

Colin Low

Hewlett Packard Laboratories,
Filton Road, Bristol BS12 6QZ, UK
cal@hplb.hpl.hp.com

Abstract

The spectre of low-cost, real-time voice communication over the Internet has polarised Internet service providers and telephone network operators, at the expense of finding solutions to the problem of integrating communications services (in particular, existing wireless and wireline telephony) with WWW content services. This paper argues that solutions not only exist and appear to be commercially viable, but could bring about a transformation of the WWW as a tool for business and personal communications.

1.0 Introduction

A red herring is an issue which appears to be important and distracts a discussion or debate away from a topic. It derives from the practice of using a strong-smelling smoked herring (or *kipper*) to distract hounds from their prey.

There are approximately 600 million wireline telephones in the world. The increase in the number of new mobile telephony subscribers alone in the past few years has matched the rate of growth of Internet users, and in gross numbers and social impact is an equally significant phenomenon - the number of mobile subscribers has reached 50 million and is expected to reach 100-150 million by the year 2000.

When the Internet and telephones are discussed together it is generally in the context of Internet Telephony, the use of the Internet as a bearer channel for digitised voice communications. Internet telephony is an important and exciting new technology which will drive many new applications and services, but the commercial value of calls placed on this medium will be insignificant for years to come by comparison with current call and service revenues generated by conventional telephony. In addition, Internet Telephony runs counter to consumer trends in telephony, which favour lightweight, pocket-sized, highly portable handsets which can be used from anywhere. In this context the debate generated by Internet Telephony is a red herring because it begs the question 'what do we do with the other 700 million telephones'.

The issue addressed in this paper is the integration of communication services with WWW information content services. In particular, it addresses the problem of integrating the vast infrastructure of public telephony, normally referred to as the Public Switched Telephone Network (PSTN) with the WWW.

The growth in WWW sites as personal or organisational points-of-contact shows there is a strong demand for integrated communications, and this is mirrored by parallel developments in the telephone industry, which sees itself as the source of communication services. This paper shows that it is possible to integrate wireless and wireline telephony with the WWW by using technology which has existed for some time in the PSTN.

The technology described within provides:

- integration of telephony services with WWW content provision.
- open telephony service development on the WWW.
- a devolved service architecture, where services are developed by users or any third party developer.
- evolutionary "market driven" development of services which parallels the Darwinian evolution of content services on the WWW.
- integration with traditional telephony services as exemplified in Intelligent Networking by telephony service providers.

The paper begins by reviewing progress in computer-telephony integration which has taken place in the PSTN over the past two decades, under the guise of Intelligent Networking (IN) [1] [3] [4] and (in the US) Advanced Intelligent Networks (AIN) [2]. It goes on to discuss parallel developments in both the Telecoms and the Internet communities which lead in the direction of integrated personal communications, and concludes that there is a great deal of value in both approaches. I show that there is a hybrid between the two approaches which provides a powerful integration between WWW content and existing telephony services. Lastly, I outline a prototype constructed by the Nexus project in HP Laboratories which provides a working demonstration of this approach.

2.0 Telephony Services - Intelligent Networking

It is not a secret that the Telecoms industry plans to replace the Internet and the WWW with a higher quality, integrated service, multi-media network which offers all the things (it is claimed) the Internet does not - multi-media communication services, managed bandwidth, quality of service guarantees, security, management, reliability, and a sensible tariff structure. An example of a developing architecture can be found in the work of the TINA-C consortium [5].

Whether these plans are based on misconceptions about the Internet, and whether these plans will reach fulfillment, will be touched upon only in passing, as it is not my intention to take sides in the polemic between the Internet and Telecoms communities. These plans existed before the Internet explosion, and they continue to develop in detail. A key technology in these plans is Intelligent Networking (IN).

0-7803-3336-5/96 $5.00 © 1996 IEEE

IN emerged in rudimentary form in the '70s. Important developments were the introduction in the telephone network of stored-program control switches as a replacement for the older electromechanical switches, and the deployment of out-of-band signalling in the trunk network as a replacement for the practice of using the voice bearer channel for signalling. The separation of voice bearer channel from signalling was an important development, and the signalling function is now implemented by the Common Channel Signalling System #7 (SS7) [9]. The separation of signalling from bearer channel allocation can be seen in its most complete form in ISDN [11], where digital end-to-end signalling is provided between user terminals, and no relic of in-band signalling (e.g. DTMF tones) remains.

The emergence of stored program control switches in the network made it possible to add additional services to switches. The best known IN services are 800 number services (Freephone), credit-card calling, virtual private networks, and a wide variety of personal call handling and redirection services (screening, follow-me, call forwarding, personal numbers etc.). The problem with adding services to switches is the complexity of switch software, the large code base (e.g. 25 million lines of code), the need to avoid compromising extremely high levels of reliability, and in consequence, extremely long development and deployment times.

The alternative was to provide an open interface in switches which could be used by external service logic to control call handling, and it is this architecture for the provision of telephony services outside of the switching or bearer network which is called *Intelligent Networking*. The main focus for IN standardisation is the ITU-T, and the Capability Set standards (CS-1, CS-2, CS-3) [12] are probably the best reference, but there is significant regional involvement via Europe's ETSI, the United States ANSI and Japan's TTC.

The principal functional components in the IN Distributed Functional Plane are shown in Figure 1.. The Service Switching Point (SSP) provides an abstraction of switch and call control resources. The Service Control Function (SCF) executes service logic external to the switch which assumes call control, making use of service, subscriber and subscription data maintained in the Service Data Function (SDF). Interaction with the subscriber, via DTMF digit collection, recorded announcements, speech recording, and voice recognition takes place via the Special Resource Function (SRF) (sometimes called an Intelligent Peripheral). Services are created using the Service Creation Function (SCF), and the remote management of service installation, subscriber data and SSP configuration is carried out using a Service Management Function (SMF). The SCF and SDF are normally co-located, and the resultant system is called a Service Control Point (SCP).

The most mature and widely supported part of IN is the SCF-SSP interface. Call progression within an SSP is represented by a Basic Call State Model (BCSM) (see ITU-T Q.1200 series standards [12]) which takes two forms, one for the originating portion of a call, and the other for the terminating portion. The BSCM contains Detection Points (DP) and call states designated Points-in-Call (PIC). The Service Management Function is used to install triggers at detection points, so that when a call passes

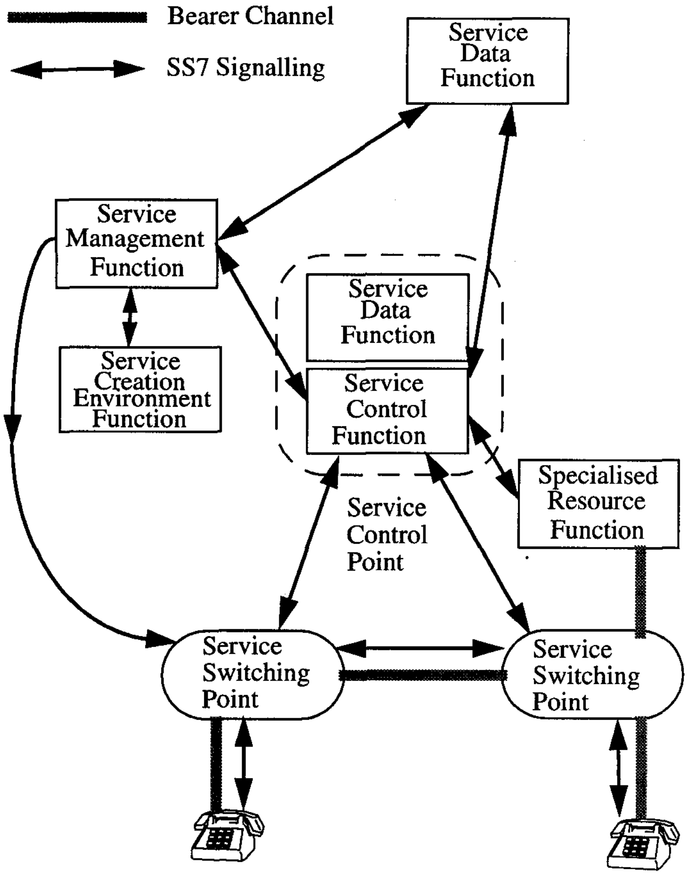

Figure 1. Intelligent Networking

through a trigger, call processing is suspended and interaction with the Service Control Function takes place, which can then take over many aspects of subsequent call processing.

Interactions between IN components are implemented using messaging over the SS7 network using the Transaction Capabilities Application Part (TCAP) [9], a variant of ISO ROSE, which provides a symmetric interaction protocol between two peer entities. The inter-component message specification is called the IN Application Protocol (INAP).

IN services are often very simple from a conceptual point of view - for example, in an 800 number service, the SSP traps the call to an 800 number, the SCF uses service logic to determine where to route the call (typically based on time-of-day, geographic factors and other customer data stored in the SDF) and the SCF instructs the SSP to dial the new number, billing the called party. Much of the complexity derives from the need to integrate new IN services with legacy billing, customer care, and management systems.

The significance of IN is that there already exists a well-developed infrastructure within the PSTN for computerised control of telephone calls. The importance of intelligence in the PSTN is increasing dramatically with the popularity of mobile telephones - mobile standards such as GSM are strongly based on ideas developed in the IN community (e.g. see [6] [7]). Telephony services are popular with subscribers, and there is now a demand for increasing levels of integration between wireless and wireline telephony, for complex personal number and mobility services, for enterprise virtual networks, for call distribution services, for multimedia content services, for video dial-tone and so on.

A problem with IN in practice is that telephony services are still deployed in the same manner as computing services in the days of the corporate mainframe, where applications were hand-crafted and operated by a large MIS department. Service Control Points are typically deployed on a national basis, carry a few services, and serve many tens of thousands of subscribers (see [13] for a typical deployment).

3.0 Content Services - the WWW

The World Wide Web (WWW) is the exemplar of a global network providing complex content services. It is open, in that anyone can connect to it and offer content services without asking permission. Internet protocols such as TCP/IP and WWW protocols such as HTTP and HTML are global de-facto standards and even public domain implementations, available at no cost, are of very usable quality. The WWW is highly scalable. Levels of performance in the public network are variable and sometimes unsatisfactory (just like the public road network), but when the same technology is deployed in a well resourced private or organisational *intranet*, performance can be excellent, and economies of scale are driving down the cost of "well resourced" to the point that high-performance WWW services are within the reach of private individuals, and service provision has become a cottage industry. Overall resilience is very high, on the "ant's nest" principle - individual WWW sites may fail, but the very large number of sites operational at any time, and the (literally) bombproof nature of the Internet means that almost all services are available almost all of the time.

There is no question that the diversity on content services on the WWW would not have evolved had the technology been available only to a small number of large, capital-rich organisations. The key to the rapid development of rich content services has been the openness and the low cost of entry, which has released a huge amount of entrepreneurial energy, a point which I will return to in the context of communication services.

Communication services on the WWW consist largely of electronic mail and form-based communication using the Common Gateway Interface (CGI). Improved compression techniques have made it possible to create Internet telephony applications, but the Internet was not designed to carry continuous media and the public portion of the Internet does it poorly at present - it is difficult to sustain voice communications on a peak hour, point-to-point bandwidth of 28 bytes/sec (observed by the author on many occasions). There is little doubt that new Internet protocols such as IPv6 [19] and the Resource Reservation Setup Protocol (RSVP) [20] will improve the situation for isochronous traffic and the long-term future for Internet telephony looks good. Bridging between Internet Telephony and the PSTN is already being provided on a commercial basis, making it possible to terminate Internet telephone calls in the PSTN.

The difficulty is that value-added communication services in the PSTN remain isolated from communication services on the Internet, creating a problem for a large population of users with Internet connectivity for whom the plain old telephone is still the *sine qua non* of day-to-day business communications. The basic aim of communication services is to improve communication by handling the exceptions - when you are out (voicemail), busy (call waiting), on another telephone (call forwarding), on the move (mobility services) and so on. Creating parallel communication services on the Internet and PSTN may have the opposite effect of making personal communication more difficult to manage (there are many more possibilities) and there is a strong argument in favour of merging the two worlds.

4.0 Two Worlds

If the WWW has a dual (in the mathematical sense), or a foil (in the dramatic sense) then it is communication services, represented by IN. IN has many strengths and many flaws. Its strength is that many services are successfully deployed, such as 800 number services, credit card calling, virtual private voice networks, voicemail, and various call waiting and redirection services.

IN's primary flaw is that despite years of standardisation, services are implemented one-at-a-time on proprietary platforms. Switch vendors (who have little commercial motivation to open up their switches) are still lagging behind the standardisation process. IN systems do not scale well, and tend to be implemented using a mainframe mentality - huge, fault-tolerant SCP systems which provide services for hundreds of thousands or even millions of subscribers and take years to deploy. The networks used to support these services also support normal telephony, so anything attached has to be rigorously vetted. Despite years of standardisation, each country tends to have local variations of international standards, making it difficult to sell standard products into this market. Given the importance of standards in the telecoms industry, this may seem surprising, but until recently, telecoms was dominated by state PTTs large enough to propose national variations, and despite the grindingly slow and thorough nature of ITU standardisation, interoperability is complex - even SS7 and TCAP, the foundations of IN, have significant ANSI variations. A major deficiency is the poor integration between fixed and mobile telephony services - the popular GSM standard for mobile telephony was founded on ideas taken from IN, but it is a closed architecture and it is possible for a person to have separate services for their domestic and mobile telephone with no way to integrate the two.

Service creation in IN is the precise dual of the WWW in that it is closed, and the cost-of-entry is very high and limited to a relatively small number of highly capitalised corporations.

Telecommunications companies believe that a next-generation architecture will support both advanced telephony services and new services such as video dial-tone, integrated fixed and mobile services, and content services such as video-on-demand (VOD), distance learning, and home shopping and banking. It is possible that the architecture-of-the-future will emerge from the TINA-C consortium, but there is also a competing claim in the telecommunications industry that IN will evolve from being an architecture for telephony services into an architecture for broadband services [14] [15]. The model is that these services will be developed and deployed in the same closed manner as IN, by a small number of highly capitalised corporations.

The issue at stake here is the future of integrated content and communications services, the point of convergence for both IN and the WWW. Will these services be provided in the manner of WWW content services - open, highly decentralised, hosted by low-cost commodity platforms, and subject to intense entrepreneurial competition and Darwinian evolution - or will they be provided in the manner of IN services - closed, centralised, with capital-intensive bespoke implementations and a slow rate of deployment and evolution.

A belief underlying the work described in this paper is that the IN industry has to embrace the service creation philosophy of the WWW in the long term. There is already a recognition that customers are looking for better service integration, for more complex personal mobility services, for easier service configuration and customisation - the explosion of activity on the WWW is a demonstration of how far customers are prepared to go in personalising organisational and personal WWW sites, and how far the IN industry is from achieving these goals. The level of interest in Internet telephony will add further pressure to bring the WWW and PSTN closer together.

The solution presented in this document for a converged content and communication service architecture, appropriately called *WebIN,* is to take the service architecture of the WWW and combine it with the control architecture from IN.

5.0 *WebIN* Physical Architecture

5.1 Outline

WebIN[1] integrates these converging technologies by identifying the Service Data Function in IN with a personal or organisational virtual presence (i.e. WWW. home pages). The SDF in IN terminology, the repository of communication service subscriber and subscription information, is removed from the closed SS7 network and relocated and physically distributed on the open WWW, where it becomes identical with a WWW site. In practice, a person can be identified *both* within the telephone network[2] *and* on the WWW by a URL or URN. The physical architecture for WebIN is shown below in Figure 2., which shows the Service Control and Switching functions in the PSTN communicating with a Service Data Function physically distributed across WWW servers.

The link between the worlds of telephony and the WWW is an Address Translation Database linked to the SCF which can translate telephone numbers into URLs. This provides the link between a telephone number (which could be a portable personal number) and personal service data in a geographically dispersed SDF. An implementation of this database using the Internet Domain Name System is briefly described below. This association between telephone numbers and URLs means that you can obtain your telephone services *even when you are not on the WWW.* The combination of an SCF with a WWW compliant SDF and an Address Translation Database we term a *WebSCP* (see also below in Figure 3.)

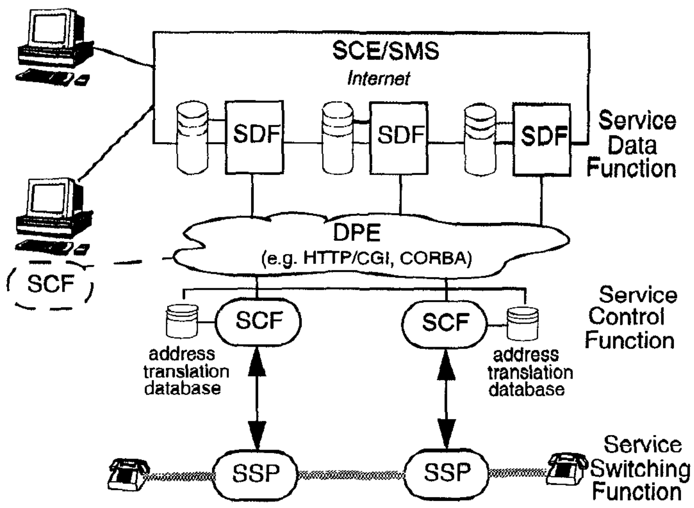

Figure 2. WebIN Physical Architecture

The architecture in Figure 2. can be viewed in two ways: as an extension to the WWW which permits WWW services to control telephony, or as a modification to IN where service creation, service data, and much of service management have been moved from a closed network (SS7) to an open network. I believe it is important to hold on to both interpretations, because they have very different commercial implications, and throughout the paper I will retain this dual IN/WWW perspective, switching between the two viewpoints, even though only one technology is being described.

5.2 Communication Scenarios

The types of communication envisaged are shown below in Figure 3. Users A and B are assumed to have connectivity to both a signalling network (Internet, enterprise intranet) and a bearer network (PSTN, ATM ...). Users C and D have connectivity only with the bearer network, in this case the PSTN.

Users A and B are shown with two terminals each, one for Internet and WWW access, the other for telephony. They could be integrated (e.g. via TAPI), or, as is still the case for most people, they will not be. Various WWW servers are shown, acting as repositories for services and objects owned by User A and User B *and* Users C and D.

Some of the kinds of interaction which are possible are:

- User A uses WWW page interaction to set up a PSTN telephone call between User A and any of Users B, C or D.
- User C makes a PSTN telephone call to User D, and both Users C and D trigger IN-style telephony services hosted on WWW servers.
- User B configures his service subscription using WWW interaction (we could imagine that User D is in fact User B when he is using his mobile telephone).

1. I am grateful to Nicolas Bouthors, Nicolas Raguideau and David Skov for suggesting this name.
2. This applies also to future switched technologies such as ATM, but for simplicity I refer to the telephone network only.

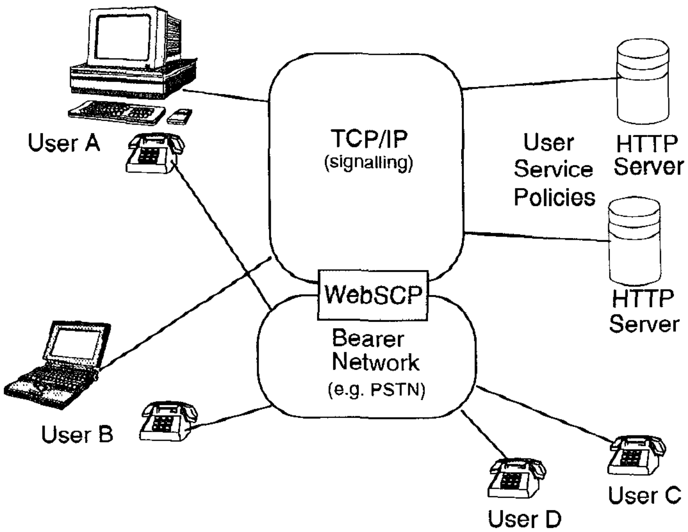

Figure 3. WebIN Interactions

- User A uses an Internet telephony application, bridged onto the PSTN, to talk to User C.
- User C dials User A's Internet telephony application *as if it was a normal telephone*.

Contact policy, that is, how a person wants contact attempts to be handled, can be unified because in this view contact services have been disassociated from a particular contact technology (e.g. telephony), and the WWW becomes the unifying mechanism which integrates contact policy for a number of different technologies. The examples given above show three contact technologies - WWW pages, Internet telephony, and PSTN terminals, but clearly other technologies such as FAX, voicemail and email can be integrated in the same way.

5.3 The WebSCP

The WebSCP unites two worlds: the bearer channel world (voice telephony, video telephony) and the WWW. The WebSCP can do four things:

- it can create bearer channel connections between terminals.
- it provides a User-to-Network signalling interface into the PSTN or a private campus network, which allows Internet telephony applications to initiate or receive calls from the PSTN.
- it can translate telephone numbers into URLs (possibly via URNs).
- it can retrieve and execute service logic from WWW servers using HTTP.

What makes the WebSCP different from a normal SCP (as commonly implemented) is that service logic is not concentrated on a closed database - it is distributed across normal WWW servers and is maintained by users in the same way as they maintain their normal WWW pages. Service logic (which could be written in JAVA) can be embedded in WWW pages (or loaded from WWW sites as JAVA classes). Control is devolved to the users, who are free to construct telephony services of arbitrary complexity and integrate these with their existing content services.

5.4 The Service Data Function

The WebIN SDF is structured as shown below in Figure 4. A

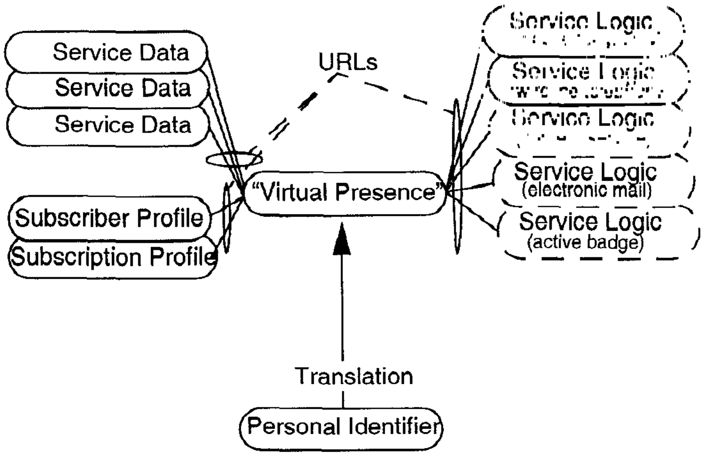

Figure 4. The Service Data Function

Uniform Resource Locator specifies an individual's communication service resources, and the resources themselves are distributed in the same way as an HTML page can seamlessly link geographically dispersed documents.

A personal identifier in the telephony world (e.g. a 700 number [10]) is associated with a URL in the WWW world. The key to the scaling, resilience, location transparency and ubiquitous access properties of the WebIN SDF is a fast, resilient, federated translation database attached to the SCF. An example of such a technology is the Internet's Domain Name Service (RFCs 1034, 1035), and we have used an unmodified, public domain implementation of DNS (BIND) to host a scalable telephone number-to-URL translation service. The hierarchical nature of telephone numbering plans lead to similar scaling properties in the database as in Internet Domain Names. In order not to impact the Internet with tens of millions more DNS entries, I assume the address translation database for telephone numbers would be distinct and maintained by telephony service providers.

The ability of the SCF to carry out a personal identifier to subscriber resource translation means that subscriber resources are disassociated from a specific SCF and a specific service, and it becomes possible to amalgamate personal service resources for several services (even across several service providers) by using Uniform Resource Locators, either in their current WWW instantiation or in some evolutionary form.

Service composition is not limited to wireline or mobile telephony - content services, email services and location services (e.g. via an active badge) can become the basis for integrated services. It is possible to integrate services across several service providers. I believe that the problems of authentication and security posed by open access to service resources (this is of concern in telephony, where the closed nature of the signalling network is not entirely a bad thing) can be approached incrementally and are ultimately surmountable - this is addressed below. In the first instance, enhanced subscriber access for service configuration is certainly achievable.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase®
Smarter legal research.