

The Future of Presence: Don't Strangle the Killer App

Presence is the killer app of the next generation of software technology. Of course, this is somewhat inaccurate, since presence – the information served up by instant messaging about the online status of users – is not an application, per se. Still, the sentiment is right. It's not text messaging that has pushed the enormous growth of instant messaging: it's presence.

Ubiquitous Presence...

Presence means many things, and is used in many different contexts, and soon will be ubiquitous – inserted in every enterprise application, pulled from every computing device, and is likely to link together an increasingly diverse set of 'users.' The association of presence with people has been the principal area of interest to date, but in the near term we will witness an explosion of uses for presence. Some examples:

'Bots – The use of quasi-intelligent agents, or 'bots, as a conversational interface to enterprise applications or enterprise information sources is growing. As a direct consequence, 'bots inhabit the buddy list, and their presence and availability is served up just like human users.

Devices – Computers and other networked devices are being brought into the presence networks within large companies. Computing-related equipment like printers, routers, scanners, and even non-computing hardware, like air conditioners, elevators, and surveillance cameras are being connected to the presence network, and serving up their presence and availability through the same instant messaging metaphor. This allows users to check the status of a printer, for example, and could direct the printer to retry a print job through a 'bot interface. Likewise, these devices could notify individuals or groups of problems when encountered, in real time.

Inventory – The development of low-cost radio transmitters that can be attached to products and parts in supply-chain inventories or in transit offer an interesting mechanism to track location and status in true real time. (Note that the bar-code scanning approach – such as that used by UPS and Fedex for packages in transit – is not even close to real time. What you get when querying their web site tools for package tracking is the time and location of the last bar code swipe, not the actual location of the package.) The recent announcement by German retail giant Metro AG about their Future Store – which relies on RFID presence detection of packaged goods in place of bar code scanning – represents a consumer market equivalent with enormous growth opportunities

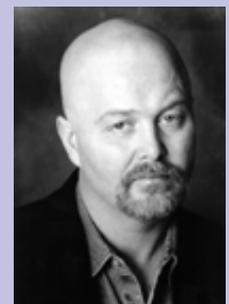
Location – Cell phones constantly update their location to their cell system, and this information is being associated with presence by many cell providers as an aspect of their instant messaging solutions. This allows individuals to be able to geographically locate their buddies when presence information is associated with maps. While the cell providers are generally touting these capabilities as something attractive to young customers – in the 15-25 year old segment – I see this as an incredibly attractive business tool, as well. Consider the use in a large corporate campus, or in a factory.

Places – In the corporate setting, it certainly would be beneficial to associate the status of physical locations – such as meeting rooms – with presence information. But the possibilities for physical locales publishing their presence are really unbounded.

Presence is the killer app of the next generation of software technology, and presence will explode in the next few years: not just the doubling or trebling of people using IM systems, but the trillions of devices, hardware, and software that will be publishing presence. But if this growth is not to be choked off, new technologies are needed.

I profile **TimesTen's** real-time event processing system, an example of the software infrastructure needed.

Editor & Publisher
Stowe Boyd



What information would be served up about groups through presence? The first would be the number and presence of its members, but other attributes might be just as interesting, like the status of the group's activities, associated documents, timeline of events, next steps, and the like.

If a group exists for a long time, the history of the group's chat is potentially an enormously valuable asset: a knowledge base to be drawn upon.

Imagine finding out where the closest empty parking space is through interaction with the parking garage's 'bot, or how busy your favorite restaurant is right now, or a snap shot of the number of cars on your commute home. By associating presence and availability information with these places, a simple and consistent interface can be exploited.

Groups and Projects – Obviously, people form groups as a matter of course, but the current notion of instant messaging buddy lists fall far short of what is needed for enterprise use. We can expect the automatic generation of groups from various sorts of analysis. Consider these examples:

- creating a small, on-the-fly team within a manufacturing company (including suppliers) to resolve a supply-chain bottleneck based on an analysis of availability, responsibility, and expertise of potential team members (as I described in *Message 3(2)*);
- contrast that with the creation and maintenance of large, long-term, long-lived communities of interest, such as the tens of thousands of foreign exchange traders in the financial services community (as I discussed in *Message 4(1)*).

These two examples form two ends of a spectrum of online groups, the first potentially existing for a few minutes or hours, and the second persisting for decades.

Work gets done in groups, so aggregating individuals into groups is obvious; however, a group is more than just a collection of buddies. Groups should have their own identity, so that the current status and availability of the group can be accessed, as well. What information would be served up about groups through presence? The first would be the number and presence of its members, but other attributes might be just as interesting, like the status of the group's activities (has the supply chain bottleneck been resolved?), associated documents, timeline of events, next steps, and the like.

If a group exists for a long time – like the foreign exchange group – the history of the group's chat is potentially an enormously valuable asset: a knowledge base to be drawn upon. While it may seem strange to consider the availability of information – the sort of thing managed in FAQs or indexed information stores – accessed through specialized presence and availability presentation, I anticipate that this will emerge. In effect, this is rendering the hits resulting from a query on the information (perhaps through a 'bot interface, but perhaps not) – in this case, the decades of stored chat associated with the foreign exchange group – as being 'available'.

I put "available" in quotes above because it's a stretch to say that the availability of information in this sense is really the same sort of thing as me being available to talk to you. However, the ease of use and simplicity of interaction in this way will make such an approach attractive.

...And its Costs

Presence and Privacy – When discussing ubiquitous presence, personal privacy immediately comes to the forefront. Do you want everyone in your company to know how much time you are spending at Starbucks? Clearly, there has to be a balance between corporate need and personal privacy. The rise of ubiquitous presence will

Message is an at-least-monthly newsletter published by **A Working Model**

Editor and Publisher – Stowe Boyd

11195 Longwood Grove Drive, Reston VA 20194 US +1 703 855 0833 stoweboyd@aworkingmodel.com
www.aworkingmodel.com

create a commensurate requirement for significantly more sophisticated privacy controls, or presence filtering. The all-or-nothing, global presence and availability controls provided by today's instant messaging solutions are already inadequate. What is needed for the enterprise is a complex series of presence and availability filters:

- **Enterprise Privacy** – Information about the functioning of online devices, project status, and the coming and going of employees within corporate buildings is clearly critical to the enterprise, but sufficient access controls must be put in place to guard against the misuse of this information, both relative to enterprise confidentiality and security, and personal privacy.
- **Personal Privacy** – Subordinated to some extent by corporate needs, individuals will require a broad spectrum of highly sophisticated controls around their presence and availability. The combination of presence (who can see my presence?), availability (who am I available for and when?), location (who can see my location and when?), and device (what device am I using?) and other elements of personal status will lead very rapidly to requirements for extremely sophisticated, complicated, and constantly changing presence filters.

Presence Management – Today's presence systems are unlikely to be able to scale up to the demands of the impending profusion of people, devices, projects, groups, and every other addressable entity connected and participating with the emerging worldwide presence network.

Leaving aside the debate about peer-to-peer versus server-based instant messaging, in both cases every presence or availability change made will lead to messages being sent through the network, informing others who are subscribing through their buddy lists to that information. Even in the most efficient presence management systems, the explosion of presence information over the next few years will be enormous. Today, there are over 400 million registered users of public instant messaging solutions, and that number is likely to double or treble in the next few years, as more users come on line and existing users increase the number of devices they use for IM. As the trillions of computing devices, projects, packaged goods, refrigerators, thermostats, cameras, and other presence-producing items come online, we are likely to see serious presence bottlenecks. Slowing down the 'ping cycle' – the frequency of presence tests – is not a good idea, but there have to be some key technologies brought on line to meet this challenge.

There are two sides to the presence bottleneck: sending all the messages, and updating presence information, the latter being more complex than the former. Presence information is relatively small – even in a world where it may have to be encrypted for security purposes. But even so, the sheer volume of future presence will mean that obvious techniques for decreasing message traffic will need to be applied to presence: for example, aggregating all presence changes within one organization and passing that as a single block to an external organization, as opposed to sending thousands or tens of thousands of separate messages.

Real-Time Event Processing For Presence Management

The highest barrier to supporting large-scale presence management is the actual updating of presence changes. The use of a conventional RDBMS is simply not a good match as volumes increase. Expediting the propagation of presence updates when the presence information is held in memory is much faster and scales more efficiently than the conventional database approach, which is geared around backing up all committed transactions to a persistent, disk-based store. This makes sense when

Today, there are over 400 million registered users of public instant messaging solutions, and that number is likely to double or treble in the next few years, as more users come on line and existing users increase the number of devices they use for IM. As the trillions of computing devices, projects, packaged goods, refrigerators, thermostats, cameras, and other presence producing items come online, we are likely to see serious presence bottlenecks. Slowing down the 'ping cycle' – the frequency of presence tests – is not a good idea, but there have to be some key technologies brought on line to meet this challenge.

The highest barrier to supporting large-scale presence management is the actual updating of presence changes. The use of a conventional RDBMS is simply not a good match as volumes increase.

TimesTen's real-time event processing system can process tens of thousands of transactions per second, an order of magnitude increase over a conventional RDBMS. TimesTen's ability to capture and process presence updates in real time translates into, for example, a commuter getting an update of traffic information in a few seconds instead of a minute, or a critical device in a nuclear power plant updating its availability in under 1 second instead of a minute.

TimesTen, Inc

John Trembley
(650) 526-7146
trembley@timesten.com
www.timesten.com
1991 Landings Drive
Mountain View, CA 94043
US

you require full recoverability and a record of every transaction, but presence information has these qualities:

- **It's transient** – If the presence server resets for some reason, it's not critical to rewind to the last known presence status for users. It's better to just restore presence based on getting the current status.
- **It's ephemeral** – In general, it's not necessary to record the changes in presence persistently. While a company might want to log all message content for audit or knowledge management purposes, it may prove to be best for privacy reasons to explicitly not capture all presence data. In specific circumstances, this may not be the case, such as monitoring user's location in secure facilities, but as a general rule, presence is – and may be should remain – ephemeral.
- **It's huge** – The only areas we can look to for guidance about the complexity explosion around presence are other experiences in large-scale real-time event processing, like financial services and telephony, where large organizations are routinely performing tens or hundreds of millions of real-time transactions daily. In particular, many of the major telecom players and financial services companies have shifted to real-time event processing systems to handle their large and time-constrained transaction processing loads – either home-grown or procured from software vendors specializing in these products.

I was recently briefed by a leader in this market, **TimesTen**, about how they have applied their Real-Time Event Processing System to supporting presence, availability, and location information with various partners and customers in the telecom sector. TimesTen's real-time event processing system can process tens of thousands of transactions per second, an order of magnitude increase over a conventional RDBMS. TimesTen's ability to capture and process presence updates in real time translates into, for example, a commuter getting an update of traffic information in a few seconds instead of a minute, or a critical device in a nuclear power plant updating its availability in under 1 second instead of a minute. The magnitude of the transactional loads in such systems – alerts, call routing, presence updates, preferences, device info, and location – will stand as the primary challenge for the real-time economy.

This infrastructure software provides both speed and reliability features which enables a blended environment where some of the transactions are safeguarded while others are not. This provides optimum performance without sacrificing the reliability of the application. TimesTen's stellar customer list – **Cisco, Alcatel, Deutsche Bank, Ericsson, Lehman Brothers, HP, Lucent**, and many others, is an indicator of the criticality of real-time event processing today, and the work they are doing with companies like **dynamicsoft** and **Teltier** (leading players in the SIP/SIMPLE and PAM standard efforts, respectively) suggests a future role in the future of advanced presence management solutions.

Close

The presence explosion is just around the bend, and even today the largest providers of presence services – like public networks, telecom players, and even large companies with hundred's of thousands of IM users are running up against the computing barriers to exploiting presence effectively. The expansion of presence to include more devices, equipment, and groups will only increase these problems. To avoid strangling presence, the killer app, IM providers and users will have to consider more advanced solutions for presence management than those currently deployed – such as TimesTen's real-time event processing system for presence management.