# *Role-Based Policies to Control Shared Application Views*

**Lior Berry, Lyn Bartram and Kellogg S. Booth**

Department of Computer Science
University of British Columbia

{berry,lyn,ksbooth}@cs.ubc.ca

The work reported in this technical report has been submitted for publication and will appear in more polished form at a later date.

## Abstract

Collaboration often relies on all group members having a shared view of a single-user application. A common situation is a single active presenter sharing a live view of her workstation screen with a passive audience, using simple hardware-based video signal projection onto a large screen or simple bitmap-based sharing protocols. This offers simplicity and some advantages over more sophisticated software-based replication solutions, but everyone has the exact same view of the application. This conflicts with the presenter's need to keep some information and interaction details private. It also fails to recognize the needs of the passive audience, who may struggle to follow the presentation because of verbosity, display clutter or insufficient
familiarity with the application.

Views that cater to the different roles of the presenter and the audience can be provided by custom solutions, but these tend to be bound to a particular application. In this paper we describe a general technique and implementation details of a prototype system that allows standardized role-specific views of existing single-user applications and permits additional customization that is application-specific with no change to the application source code. Role-based policies control manipulation and display of shared windows and image buffers produced by the application, providing proactive privacy protection and relaxed verbosity to meet both presenter and audience needs.

## *Introduction*

People working in groups increasingly rely on the ability to share views of an entire work session or a specific application for co-located or distributed cooperative work. Although new tools and frameworks introduced in recent years support a wide range of collaboration formats, the dominant format is still that of a single person, the *presenter*, sharing a view of her workstation while others, the *audience*, watch. This *generalized presentation* setting applies to people giving conference or classroom presentations, demonstrating software, training others, or engaging in collaborative work where a shared document on a public display is the focus for group discussion. The same pattern is applicable during turn-taking editing sessions where at any given time there is only a single editor (the presenter) and all the others are just viewers (the audience).

For co-located groups, view sharing is often done by replicating the video signal from the presenter's computer onto an external public display. For distributed groups bitmap-based screen sharing protocols such as VNC [11] provide view sharing. Functionally, both solutions are equivalent and afford three properties that more sophisticated collaboration-aware solutions may be hard-pressed to achieve: (i) collaboration is transparent because any application can be shared without requiring it to know it is shared; (ii) viewers do not need to install the application, which is often an unacceptable imposition due to licensing or other issues; and (iii) there is no need to synchronize views between the presenter and audience because only one copy of the application is running.

The last point is an important benefit, but it is also the chief drawback of these schemes. A strict ``What You See Is What I See" (WYSIWIS) mode is imposed -- all viewers are forced to share the exact same view that the presenter sees, despite the roles that they play within the group. This conflicts with the different needs of the presenter and audience, so a ``relaxed WYSIWIS" mode has been suggested in the literature [13]. Unfortunately, none of the current bitmap-based solutions address this successfully. More elaborate collaboration-aware solutions can and often do present different views to each participant but only at the expense of application-specific features and coping with complex synchronization [16].

## Motivation

The initial problem motivating our work was the lack of support for presenter privacy. While generally interested in sharing a view with her audience, there are often interactions or document components a presenter would like to keep private. These may be interactions with other running applications on the desktop (such as a text editor with notes) or with parts of the shared application that are deemed private (e.g. a file open dialog listing documents in a private folder). It is clear that someone engaged in a live presentation might not attend effectively to her privacy at the same time. Moreover, as public displays and the means to drive them (such as dual-video-card laptops) become ubiquitous, shared presentations will become part of day-to-day work, so presenters may find themselves in a presentation-like mode without a chance to prepare or while having to do other tasks on their computer while ``on-the-air." These make privacy a problem. The need to limit the information that is shared on the public display cannot be ignored.

A quite different problem arises when we consider the audience experience. Passive viewers may wish to control the type and level of information presented to them and they may require assistive cues to accurately follow the presenter's interactions with an application. Passive viewers frequently find themselves searching for the current point of interaction (a problem intensified on large screen displays [2] or being forced to view tedious interactions by the presenter (such as searching for a menu item or adjusting display parameters) that are irrelevant to their interests. So we may need to display less information, or we may need to display more information to compensate. Some needs are already addressed by screen recording tools, like Camtasia.[1] Some visual enhancements and cleanup can be applied to a recorded movie in a separate editing session. Being able to apply these enhancements in real time is beneficial for effective collaboration.

These two problems are actually both manifestations of the need to provide role-appropriate views of an application to each group member. It is possible to use collaboration-aware tools or to adapt existing tools to run in a synchronized mode, so that presenter and audience have similar but non-identical views. While these custom solutions enable one to flexibly craft views as desired, they fall short of meeting the advantages of existing bitmap-based protocols outlined before. The proof of this claim is the fact that time and time again people resort to bitmap-based sharing rather than more

[1] http://www.techsmith.com

sophisticated solutions. Thus, any improvements to bitmap-based sharing modes are still very relevant to improving collaboration.

### Our contribution

We have developed a novel proactive framework for adapting the live shared view of applications to meet the presenter's privacy requirements and to provide viewers with suitable cues and level of detail, balancing concerns for privacy and awareness. Our system uses bitmap-based techniques to transparently share visual information, while allowing policies to be specified that control the generation of different views for the different roles within a collaborating group, by reusing the visuals from the running application.

The system conducts an ``over the shoulder'' monitoring of what the presenter is doing, actively manipulating the published visuals in three ways: (i) *spatial* and hierarchical transformations for selective sharing, repositioning and scaling of application components (including sub-window regions); (ii) simple local or global *"chromatic"* image filters, such as blurring, applied to the visual surface of the application; and (iii) *temporal* or application-state-based transformations applied to the timeline of captured interactions.

To make these manipulations more useful, some reliance on application-specific semantics is required to extract locations of semantic UI objects and to acquire the application's state. We designed a plug-in architecture and used a set of heuristics for obtaining such semantics without giving up too much generalizability.

We implemented a prototype of the system and demonstrated how it can be applied to several commercial off-the-shelf popular applications, disproving to some extent the misconception that bitmap-based application sharing forces strict WYSIWIS (see discussion in [3]).

## *A Sample Scenario*

Bob and Carol are both managers and Ted and Alice are team members in a group of employees. There is a briefing in which a single presenter, Bob, is reviewing with Ted and Alice the team's budget using a spreadsheet on Bob's laptop that is being projected onto a large shared screen viewable by Bob, Ted, and Alice. Carol is logging into the meeting remotely, using VNC. Bob is making modifications to the budget on-the-fly during the meeting.

Some of the data and parameters in the spreadsheet are confidential and should not be exposed to Ted and Alice, but should be available to both Bob and Carol (see Figure 1). Some parameters are in separate worksheets that should not be exposed by Bob, others are located on worksheets he needs to share with everyone, and he needs to resolve a few of them using other applications or files (e.g. an IM client with Carol). A key requirement

is that Bob must see the private parameters and change them during the meeting, but without exposing them to Ted and Alice.

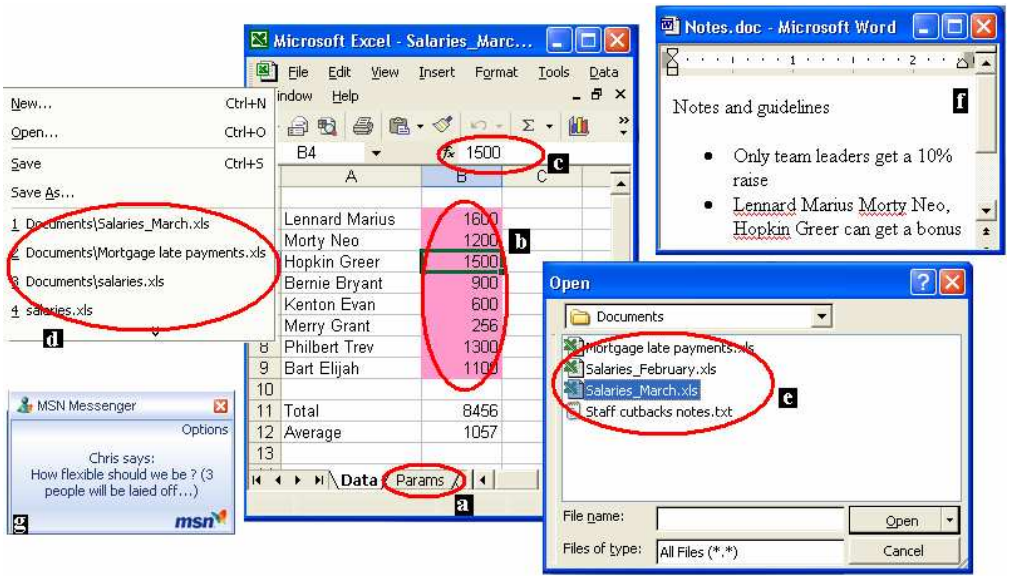


*Figure 1: Bob's privacy needs: The ``Params'' worksheet (a) is entirely private, the range of cells with individual salaries (b and c) is also private, and the file dialog and file menu (f and g) expose private files. Bob uses the IM client with Carol.*

Bob could extract just the relevant data to a new spreadsheet and project only that on a shared auxiliary screen with the master spreadsheet and the IM client visible solely on his laptop screen (also viewed by Carol). In theory this solves the problem, but not in practice. Extracting the appropriate information with its dependencies may not be trivial, but certainly it is possible. However, synchronizing the spreadsheet versions when significant changes are made is time-consuming, error-prone (even when using automated scripts) and requires redundant computations. Bob still needs to make sure that updates do not reveal private information to Ted and Alice and at the same time must keep Carol updated on the private version.

Both Bob's cognitive overhead in managing the session and Carol's limited communication channel diminish their ability to focus on the budget. Carol has to watch verbose UI interactions Bob makes, some of which are distracting (like his erroneous interactions with a wizard) and they take up her screen space. Ted and Alice, on the other hand, get to see everything that is intended for them, but looking only at the auxiliary screen they do not see any interaction cues such as menus or cursor motions; they only see the results of Bob's manipulations  echoed in the secondary spreadsheet.

This scenario demonstrates the need for role-based viewing policies. Bob and Carol need to see a different view than Ted and Alice, because of privacy concerns. All passive viewers need an augmented, ``cleaned up'' version of what Bob sees in order to comfortably follow his actions.

## *Privacy Concerns*

In the context of live generalized presentation scenarios, we make a distinction between two types of privacy concerns. There are privacy concerns that effect security (e.g. revealing one's credit card number), and there are privacy concerns that do not pose a real security threat, but may put one in an awkward position. As Palen and Dourish [10] note ``... in video conferencing, shared calendar management and instant messaging communications, concerns most salient to users include minimizing embarrassment, protecting turf (territoriality) and staying in control of one's time." Our system supports both types of privacy concerns.

We identify four categories of private information.

***Task-specific*** - A presenter needs only to expose the windows and components that are part of the shared task. This may entail sharing several applications or only a subset of a single application. It is often not necessary (or desirable) to expose the entire desktop.

***Window-specific*** - An application usually comprises more than just a single window. There are dialog boxes, menus, palettes, etc. In many cases these contain private information (e.g. recent files) or appear at awkward moments (error dialog boxes).

***Content-specific*** - This is private information visible on a window's surface, such as underlying application objects or UI widgets (a specific range of cells in Bob's spreadsheet or paragraphs in a text document).

***Application-specific*** - Some states of the application can be considered private. They may be mapped to objects that are externally inaccessible, mapped to a large set of objects that cannot be treated individually or cannot be associated with a particular object. Any visible information or interactions in this state should not be exposed (for example, the activation of a private worksheet, a show-comments mode or when arbitrary error occurs).

We can distinguish between several sources of private data.

***External data*** - Data that is not part of the currently shared document or task, such as recent files or browser navigation history, are often private. There is an ever-increasing number of these information bits, as applications become more personalized and cached versions of user's preferences and selections appear without explicit action on the part of the user.

***Semantic objects*** - Part of the shared document model, these may have several visual representations. Often, a presenter would like to specifically mark these objects as confidential or private.

***Interactions*** - These may be deemed private because they affect the presenter's reflected image, regardless of the data they operate on. Some examples are committing syntax errors, searching for the right menu item, or struggling with a wizard.

We need to make distinctions based on how predictable is the exposure of a piece of private information. The exposure of some pieces of information is an immediate outcome of the presenter's direct manipulations and fits well within the presenter's mental model of the application. These may be avoided or bypassed by the presenter at the price of forcing clumsier interactions or more careful preparation ahead of time. Other exposures are byproducts of agents that work on behalf of the user (e.g. an error message, the contents of an auto-complete text widget, or the navigation history list). These are less predictable and require more automated help to avoid accidental exposure. In either case it is disclosure to viewers we need to control, not the appearance or content of these elements.

It may seem that some of the private interactions described above are very brief, and thus the amount of information viewers can extract is limited. However, it is very common for shared sessions to be recorded, allowing later analysis that makes ephemeral information persistent [10]. Even without recording, it has been shown that viewers are more likely to read sensitive text on a large-screen public display than on a smaller personal display [14].

While our focus is on controlling visibility of private elements, it is also desirable to control their modifiability or access. An extension of the system that regulates access control deals with this.

## *Improving Viewer Experience*

In our scenario Carol, Ted and Alice are passive viewers of Bob's manipulation. We can customize each person's view by adding, deleting, or modifying the application's presentation (bitmap) to provide a more useful experience.

### Inadequacy of single-user GUIs for passive viewers

Passive viewers must follow the interactions performed by the presenter, but there is a perceptual gulf between presenter and viewers. While the presenter translates her intentions and semantic-level operations into GUI interactions, the passive viewers are doing the reverse process, inferring the underlying intentions and semantics from the interactions. This is not an easy process, even when verbal explanations are provided by the presenter (these are usually insufficient, inconsistent, and they require extra effort on the presenter's part).

One of the root problems is that the visual language of most GUIs is highly tuned for a single active user, ignoring the needs of passive viewers. For example, when a presenter decides to perform a contextual menu selection, the cue for a passive viewer that

some interaction is about to take place is the appearance of the menu, by which time it already obscures most of the context for the operation (Figure 2a ). Another problem is that passive viewers tend to follow the presenter's point of interaction (often highlighted in GUIs), yet in some cases the presenter would like to draw their attention to other regions (Figure 2b).

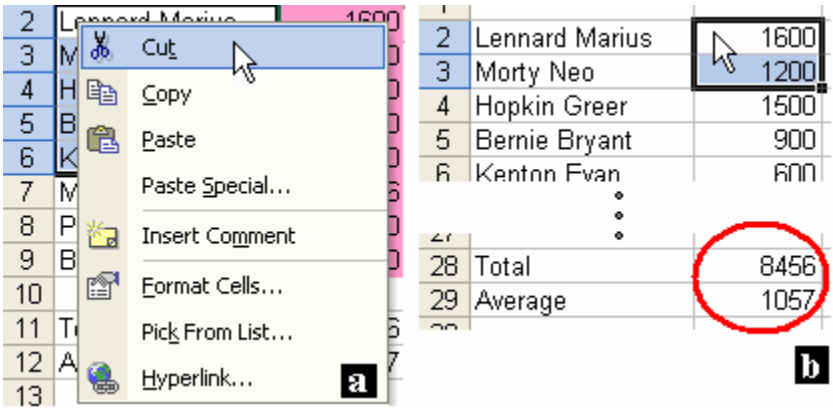


*Figure 2: Inadequacy of conventional GUI for passive viewers showing (a) context obscuration, and (b) a situation where the presenter interacts with the top cells, but wants viewers to focus on the bottom cells.*

Other low-level parameters such as cursor size or shape, the time a menu selection remains on-screen, or interaction without a specific visual indication (e.g. keyboard shortcuts), are tuned for the performance of a single active user. These are not suitable for passive viewers, who are trying to follow the interactions without the benefit of knowing the intention of the action or experiencing the kinesthetic feedback of mouse or keyboard interaction.

## Controlling verbosity

Viewers may have different levels of expertise and familiarity with a shared application. It is beneficial to adapt their view to this level. A key aspect to be controlled is the *verbosity* of interactions. For example, if Bob is to teach Ted and Alice how to fill out a report using an application unfamiliar to them, exposing the fine details of his interactions (menu selections, dialog boxes, etc.) and adding cues (like keyboard shortcuts and change highlighting) could be crucial. On the other hand if Ted and Alice are experienced users, exposing detailed interactions will prevent them from concentrating on the report semantics. From a pedagogical point of view, it sometimes makes more sense to show one logical interaction unit as a single visual step so the high-level semantics are not obscured by the low-level details.

## Mitigating visual clutter

We have already implied that it is desirable to share only relevant windows or components, rather than the entire desktop. This can assist viewers in making better use of their screen space (especially if they need to work with or view other applications

in parallel). However, if all sub-windows, dialog boxes and menus are shared as well, it can quickly clutter a viewer's display. This is somewhat like violating acoustical privacy with cellular phones [10]. The presenter imposes his “conversation” with the application on the viewers, much like a person talking on a cellular phone imposes on others in a public area. A viewer should be capable of controlling how much of this conversation penetrates his display and replace some interactions with other ``low volume'' representations. In all cases viewers need to maintain some level of awareness of the presenter's actions, but not always in a one-to-one manner.

## *System Description*

We have implemented a prototype of the system in C\# on Windows and tested it with three widespread commercial applications (MS Excel, Word, and Internet Explorer). The principles apply to any modern operating system and they work with any application, although some ``semantic glue'' layers that we describe later may be required.

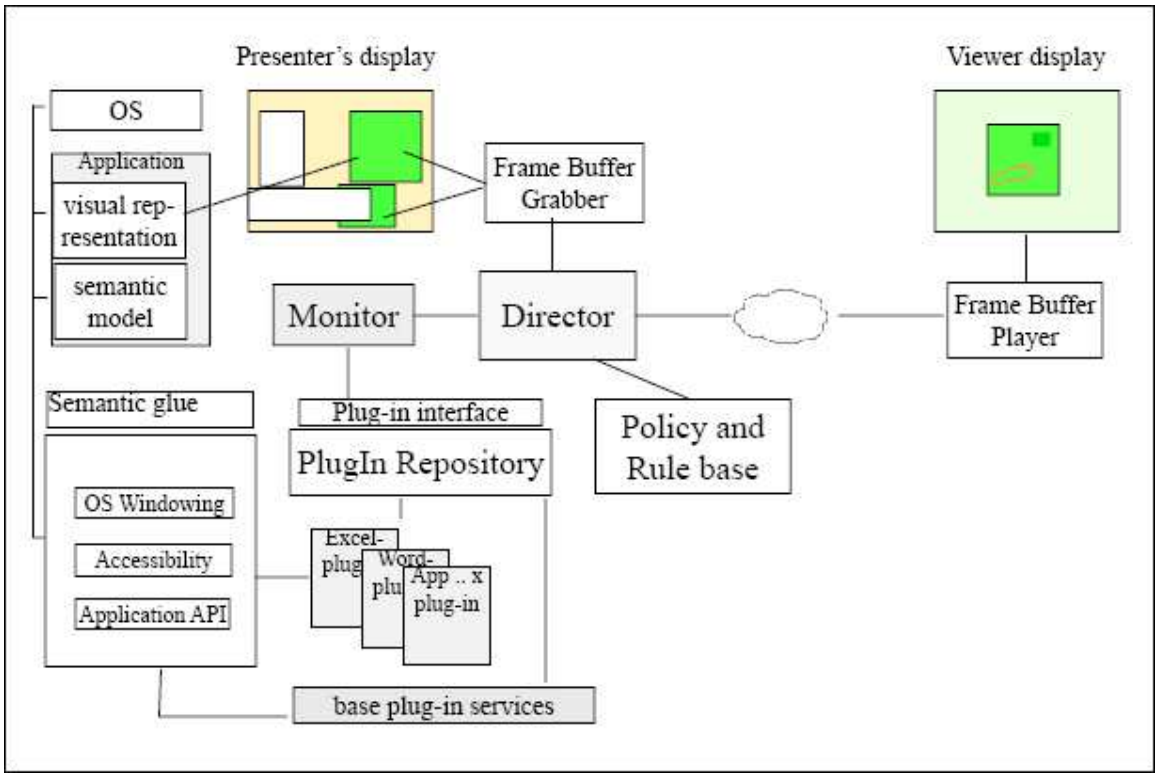


*Figure 3: System architecture. The viewer client (upper right) requires only a ``thin'' frame buffer player.*

## Cloning Windows

In order to support differentiated views, the system should first have the ability to grab the visual surface of shared application windows on the presenter's machine and convey a

manipulated version of the bitmap to a viewer's display (published in *clone windows*). In this way the viewer client can be a very thin image buffer player (Figure 3).

The process of capturing a window's image and transferring it to a different machine can be realized using a modified version of the Remote Frame Buffer protocol [11] that has to be adapted to work on separate windows. In the prototype, we used a less optimized technique, relying on timer-based copying from the window's device context similar to MSR's Wincuts [15]. This matched our initial focus on co-located scenarios (where the presenter's laptop generates both views).

Rather than a conventional video signal duplication onto the public display, we use the extended desktop mode. The public display is a continuation of the presenter's desktop (although often physically located on a wall behind the presenter). A clone of each of the shared application's windows is created by querying the system's list of windows and making bitmap copies. The clones are automatically placed on the part of the desktop lying on the public display. The presenter can move any of the application's windows on her display and cover these with other windows without affecting the published clones.[2] The novelty is in how we modify the bitmap images and window set before they are placed on the public display.

## Applying "semantic glue"

To alter the shared view along the lines discussed in previous sections, the system needs to monitor a shared application. It should be able to tell where visual representations of private elements or elements that need verbosity adjustments are on the visual surface and determine if the application is in a private state. This requires methods for obtaining information about the application's GUI components, the underlying semantic objects, and their visual representations. The following query layers are used.

***L1: OS windowing queries*** - Enumerating all windows belonging to a specific application (or process), detecting creation/destruction of such windows, visibility, titles and locations is possible in a modern operating system. Many of the widgets used in an application are themselves windows and can be accessed the same way. This layer also supports capturing of keyboard and mouse events.

***L2: Accessibility API*** - These are common APIs often targeted to sight-impaired users. They enable third-party tools, like screen readers, to systematically expose information about UI elements. We have successfully re-purposed these APIs as a resource to expose elements that should be kept private or highlighted (e.g. GUI components, menu items, rendered HTML objects). These APIs are supported by many commercial tools and UI toolkits.[3] Our use of them can be further generalized.

---

[2] see [15] for more details

[3] Supported by Microsoft Accessibility, Java APIs, OS X and more. Some level of accessibility is now required by law and will no doubt increase over time.

***L3: Application specific API*** - Many commercial applications provide an API for integration and automation. These APIs can be used via COM (Windows), JavaBeans, AppleScript and other frameworks. Within our system, we used these APIs in a very simple manner to extract information on the application's state and identify the visual representations of semantic objects. While writing some code to work with the API is required, our experience when developing the prototype shows that this is a very focused effort with a limited amount of coding. Modern APIs already provide methods for locating an object on the document surface or an Application object can usually be queried for its current state. Furthermore, coding occurs only once and can then be used in flexible ways.

***L4: Extracting information from surface drawing operations*** - This is a somewhat limited technique introduced in [9] (see the related work section). Its requirements are quite problematic, especially for the commercial tools we worked with, therefore this technique was not used and it is not shown in Figure 3. However, it is still a possible semantic glue layer that could be used in some cases.

To create a generalizable framework, we chose to provide all of these methods using a plug-in architecture for our system. Each shared application has a middleware plug-in that functions as the semantic glue. A plug-in encapsulates the knowledge about a specific application and its monitoring, and supports a common API that the Monitor module (Figure 3) can use. The PAPI (Plug-In API) provides methods that return text or keyword-based descriptions of the current state, dialog or palette. Other methods extract lists of areas of the visual surface containing private information that need highlighting or specific sub-window areas to be displayed (instead of the full window). Each of the areas are accompanied by keyword descriptions that can be used to control display parameters.

A plug-in translates these general PAPI queries into appropriate queries in one of the four layers. For example it uses L1 and L2 queries in Excel to check if the file open dialog or cell protection tab are visible and L3 Excel API calls to obtain the currently visible worksheet or comment to determine the application state. Similarly, a private regions query is translated to L3 cell locating calls or L2 queries to locate the formula auditing box and toolbars. Similar queries in Internet Explorer may result in a L2 query to check if the favorites menu is visible, L1 query to search for auto-complete text window boxes or L3 to look for private HTML fields.

Default *base plug-ins*, providing a set of application-independent capabilities to track menus, dialog boxes and other common entities, were also developed. They serve as a toolbox for developing more specifically tailored plug-ins.

The Monitor module directs its calls to a plug-in repository manager, which loads the appropriate application plug-in at run time (possibly even from remote servers). If no appropriate plug-in exists the default base plug-in will be used, offering some monitoring capabilities. (It could query the presenter before displaying any menu or dialog box, and then apply ``program by example'' techniques.)

The Director module handles the published representation to be played on the public display. It uses the Monitor module to track the application and extract descriptions of its state and visible elements. It then applies policies that determine how to manipulate the visuals.

## Policies and rules

When instantiating a policy, a tuple comprising the application, the state or element, and the viewer's role is the input. The output is a rule that determines what manipulations will be applied to the published visual representations. We must determine how private elements, states or elements that need verbosity control can be extracted, assuming shared applications do not know about privacy. There are two complementary approaches to consider.

The first approach (taken in the initial prototype) is letting the presenter mark these elements explicitly. When working with a document in an editor we can readily support what we call a *"Magic Marker"* that maps a visual property of an object to a privacy state (most editors have a notion of object style properties).

For example, a presenter can mark a document object as private by coloring it with a specific color, using the native application tools (e.g. a background color for cells in Excel or a highlight color for paragraphs in Word as shown in Figure 4). When writing in this color the semantic glue layer will recognize these objects as private. A policy that regulates blurring for marked objects will create the effect of a magic marker that cannot be seen by viewers, while the presenter can interact normally. This mode provides visual feedback and awareness on what the audience cannot see (demonstrated in [12] to be crucial).

Other means for coding attributes can also be used (like adding a ``Private'' prefix for a worksheet's name or comment text). Another option is to use the application's built-in selection mechanism, so for example the paragraph containing the insertion point can extracted by querying the application.

The situation for UI widgets is different. The generalized marking scheme we have been using relies on recording the path to the specific widget on the Accessibility widget tree or the window tree. These may be provided as a script for the plug-in. A similar scripting approach can be applied to extracting application states (in the prototype most of these are hard coded). In the general case a plug-in may expose its own UI to allow application-specific privacy control.

A second approach is to use a rule-guided search for privacy *leaks*. We have experimented with searching for private text in a spreadsheet or document (phone numbers, names, etc.) and automatically blurring them. Another interesting domain is web pages, where it is possible to search the HTML code for both private UI widgets and content elements (for example, searching for form fields that may contain private information, such as userids, e-mail addresses, etc.). It is possible to run similar searches

on dialog boxes or other UI widgets through the Accessibility API, extracting their text and determining if they convey private information (e.g. search for error messages, field names related to security, network settings or personal information) and then apply a suitable policy to issue blurring or indicate a private state.

We believe that a combination of these approaches is required for better privacy protection. Together with the visual manipulations (described later), it allows a flexible range of rules: ``do not expose any dialog related to files or the network in any application to any public viewer,'' ``blur any document element in any application marked in pink to group A members'' or ``if there is any viewer from group B, do not expose in a public view any web page not coming from company servers.''

Our prototype is a work-in-progress that serves as a proof-of-concept. We have concentrated on the system architecture and a collection of manipulations (discussed in the next section). Our goal has not been to develop a robust mechanism for describing policies, only to illustrate the value of having role-based policies to control views. This is probably the most fruitful area for future investigation.

## *Manipulating The Visual Representation*

The Director component takes in the ``raw'' captured frame buffers grabbed from the application windows and applies one or more of the following manipulations.

### Blurring

When private elements are visible, the challenge is guaranteeing viewers cannot see them, while the presenter works freely. The PAPI can extract the locations of such elements on the visual surface at any time (with attributes and hints, such as the suggested blur effect to use). In some cases a private information unit may appear in several places (e.g. the contents of a selected private spreadsheet cell will also appear in the formula bar). This demonstrates why tighter integration with application semantics is crucial for ensuring privacy.

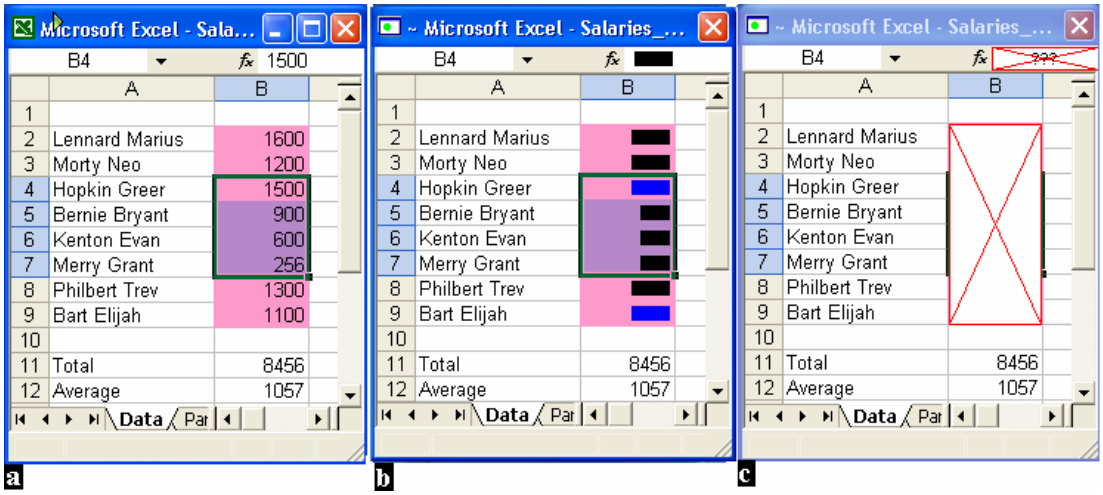

*Figure 4: (a) The presenter view of a spreadsheet, (b) Greeking cells marked with a pink ``magic pen'' exposing the presenter's interactions and orders of magnitude for the values, and (c) complete draw-over of a cell range.*

The Director can apply several image blurring operators on extracted private zones (Figures 4 and 5). Since blurring occurs at the frame buffer level it can be applied regardless of what the underlying element is (UI control, text, image etc.) or how the bitmap was drawn. Different filters offer different visual affordances, balancing between the presenter's privacy and the audience's awareness.

- ***Draw over*** - Invoked for full privacy, with no awareness.
- ***Greekify*** - Creates a ``Greeked text'' effect by searching text line boundaries on the image and replacing them with filled rectangles. Useful for exposing structure, style and some notion of the presenter's interactions (such as selection).
- ***Pixelize*** - This is a general purpose filter, mostly useful for image-based content. It provides awareness cues for viewers, but may be insufficient for privacy.

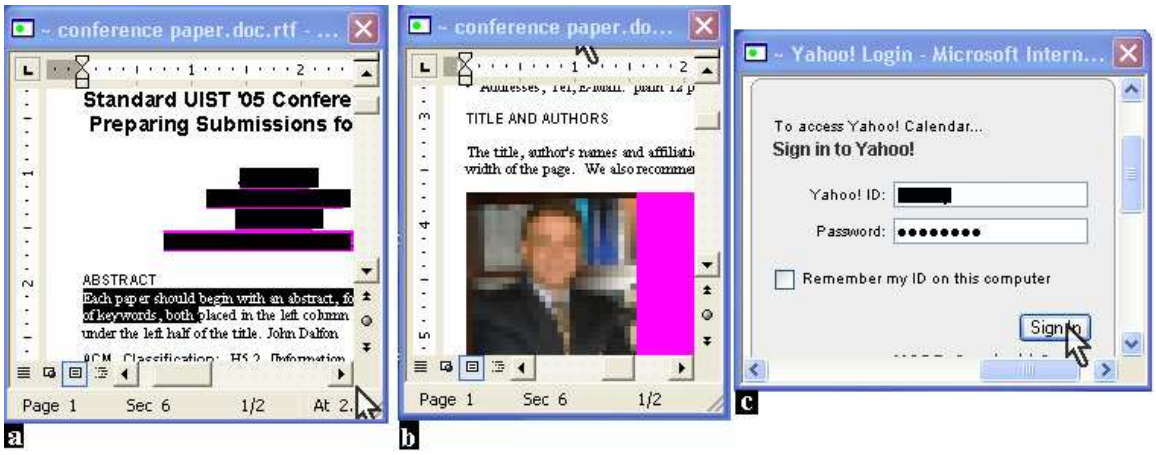


*Figure 5: A magic pink marker was used to mark private paragraphs: (a) Word document with a private paragraph automatically Greeked, (b) a paragraph containing an image pixelized, and (c) a login web page where the userid field is detected by the system and Greeked.*

## Salience and highlighting

The system supports a highlighting mechanism that is independent of the shared application's own selection and highlighting tools. In Figure 6a the presenter is interacting with the tools palette, but wants to keep viewers focused on a specific paragraph. The PAPI provides a method, through which the shared application can be queried for regions to highlight. We found it useful to highlight the context for the active selection as the contents of the document change. This is application-dependent (the paragraph, sentence or section containing the active insertion point in Word, the table surrounding the selected cells or dependent cells in Excel). The system has a unified framework to handle image buffer manipulations, regardless of application semantics. We chose to produce highlights by placing a semi-transparent colored mask on top of non-highlight areas (visually similar to [9]). Detaching the highlighted object from selection is also useful and can be done by caching the previous highlighting bounds or

by caching a pointer to the previously selected object within the plug-in (detaching / attaching is controlled via a keyboard combination). Another mode of highlighting makes changes more salient to viewers (mostly indirect changes in parts of the visual surface far from the presenter's interaction). A considerate presenter would point out these changes to an audience and perhaps even mark them on the screen. To assist the presenter, the semantic glue layer can extract such changes and provide automatically generated highlighting (Figure 6c shows hand-style circling of changed cells that also exposes changes in blurred data).

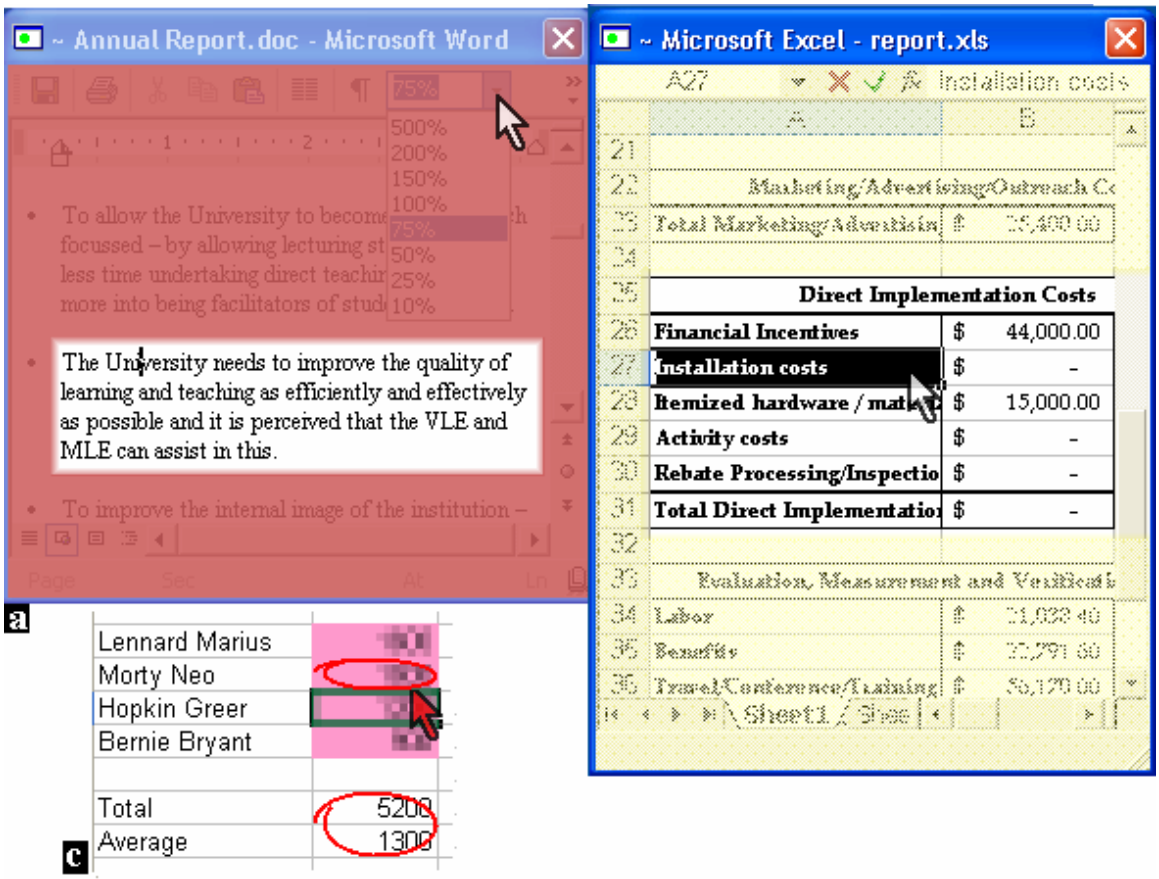


*Figure 6: Highlighting of the active context: (a) a paragraph is highlighted while the presenter interacts with other components, (b) outlining table, and (c) ``by-hand'' style circling of changed cells.*

Other manipulations that can affect salience and attention are magnifying relevant regions of the visual surface or even re-rendering of textual elements in a bigger font (many of these underlying texts can be extracted through the semantic glue layer).

## Spatial manipulations

One set of manipulations allows the presenter to share only a partial view of an application's window. This is useful for reducing screen space use and clutter, and in addressing privacy. The PAPI provides a method through which the window part to be shared can be accessed. Computing this window part can take into account several policies.

- ***Excluding the UI*** - Remove UI layers such as toolbars and embedded windows that take a substantial amount of screen space.
- ***Active context*** - Share only the active context, based on the presenter's selection as described in the previous section.
- ***Sharing a specific element*** - Sharing only a specific paragraph, table or UI element chosen by the presenter.

The semantic layer guarantees that the window part computed will adhere to the stated policy and will take into account changes to the window's dimension, scrolling or UI changes, as opposed to the manual definitions presented in~\cite{Desney2004} or to fixing a portion of the screen to be shared.

Another set of manipulations uses affine transformations. Rotating windows is very useful for single tabletop display sessions, where viewer orientation should be part of the policy. Automatically scaling down the size of dialog boxes, palettes and other secondary windows (identified by PAPI), together with a careful placement of these next to the full sized main window can assist in reducing clutter and support privacy (Figure 7d). Combining several spatial transformations together can be quite powerful. For example, it is possible to publish only the selected paragraph context, flipped vertically so a viewer on the other side of a tabletop display can follow the discussion without requiring replication of the full document window(Figure 7a).

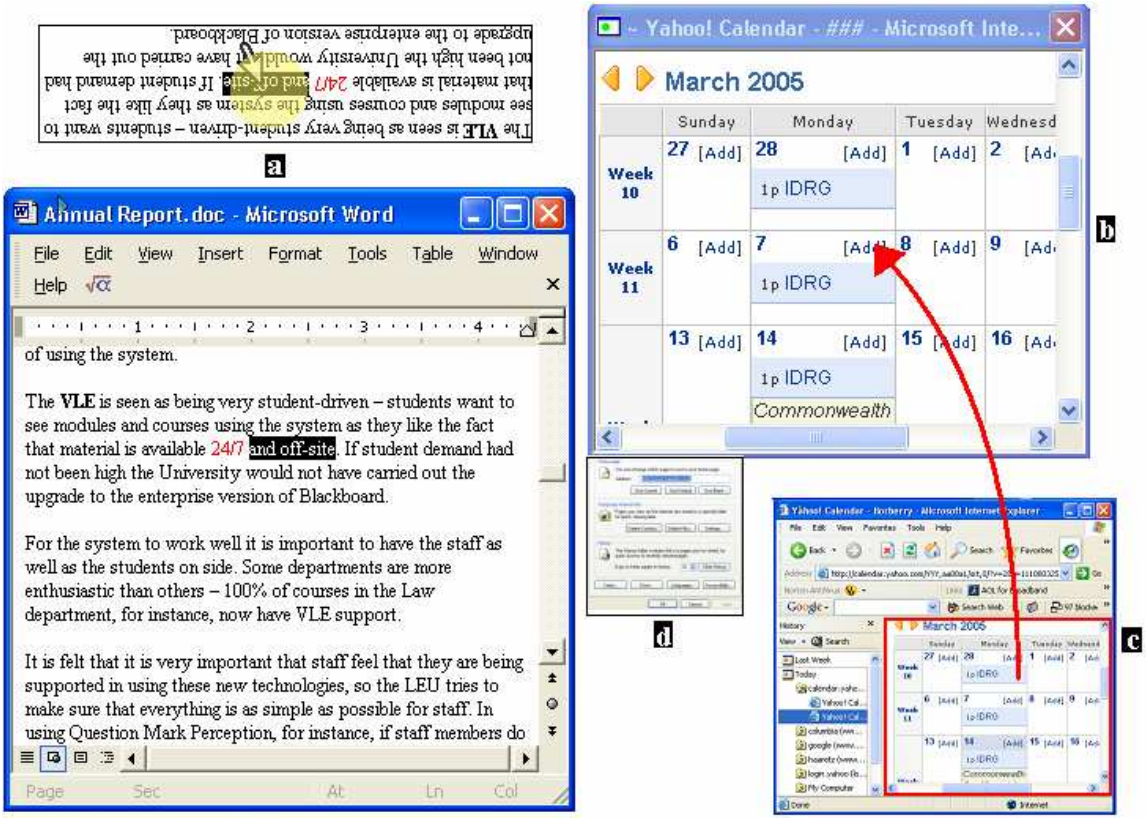

*Figure 7: Spatial manipulations: (a) publishing only active context (paragraph containing insertion point) + vertical flip for tabletop displays, (b) auto-exclusion of toolbars, menus and embedded frames from an Explorer window (compared to the size of original, and (c and d) automatically downsizing an ``options'' dialog and attaching it to the main window.*

### Temporal manipulations

There are situations when it is more reasonable to define the entire state of the application as private, rather then extensively applying blurring transformations (e.g. when the presenter interacts with a private worksheet, uses the file open dialog, or works with a wizard).

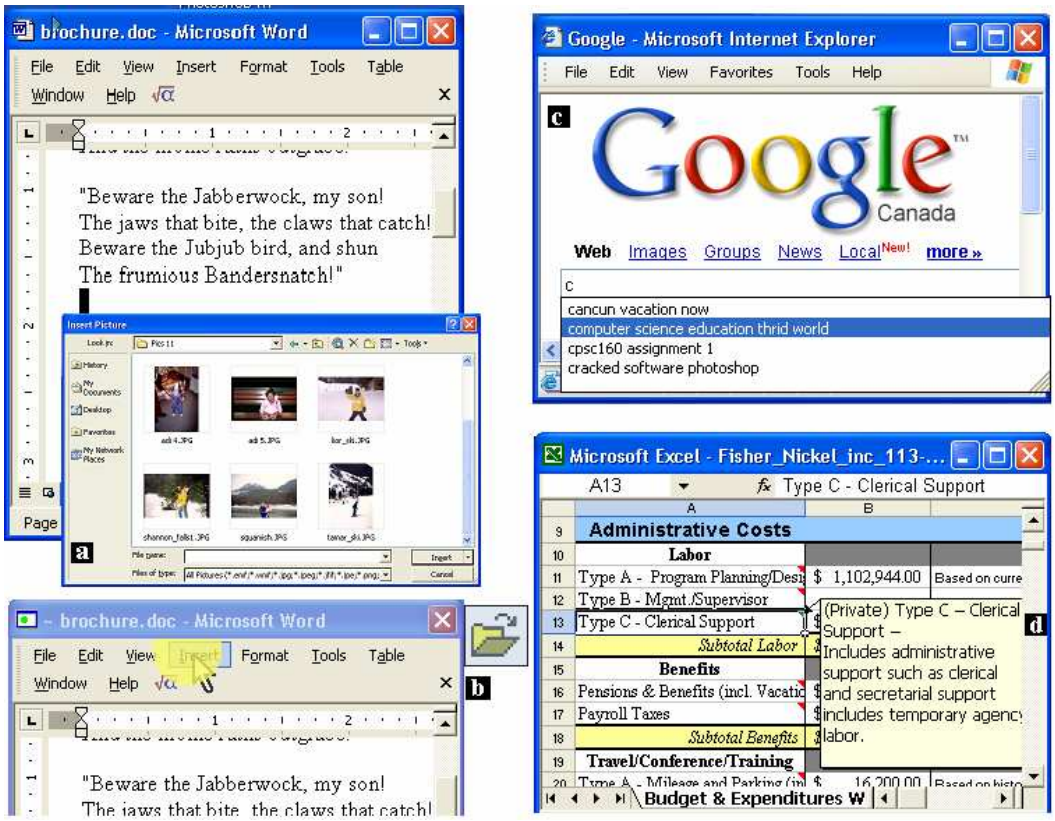


*Figure 8: File open dialog (a) is dynamically replaced with an iconic representation (b). The system will treat the auto complete text box (c) and private comment (d) as private states and will not expose them.*

The semantic glue layer can query the application state. If it matches the privacy policy, the system can trim the interaction timeline by not sending updates to the viewer's display until the presenter exits the private state. To keep some level of awareness for the viewer, the system can display an iconic representation summarizing the state. For example, an open file icon appears instead of exposing the dialog itself (Figure 8) to prepare a viewer for a document change or an icon indicating interactions on a private worksheet gives the viewer hints about what the presenter is doing.

In other situations, it is better to not provide any indication at all, maintaining complete privacy. For example, when the presenter interacts with an auto-complete text box

(Figure 8c), an error dialog, a private comment or commits syntax errors when taking notes in public. Some of these states are quite unpredictable by the presenter (e.g. an error dialog popping up or auto-complete suggestions), so automatic detection of these is crucial.

Another set of timeline manipulations can be applied to the place at which certain operations (as extracted by PAPI) are played on the public display or by letting viewers roll back recorded interactions (for instance using the semantic glue to tag recorded interactions for quick roll-back). One example we implemented involves menu selections, discussed in the next section.

## Handling Menus

Menus are fundamental interaction components that are often problematic in a generalized presentation scenario. They create a lot of clutter (being arbitrarily long, regardless of the size of their parent window), and they often bundle private information (recent files, bookmarks). More importantly, menus are becoming highly tuned for the active user and less for a passive viewer (adaptive menus with personalized order and gestural menus that do not show on the screen). It is possible to obtain a full description of a menu (items, locations, selection) through Accessibility APIs so the manipulations presented below are highly generalizable.

At the basic level, menus are windows and as such can undergo any of the previously mentioned manipulations. For example, blurring recent file menu items or marking the exposure of the favorite links menu as a private state. A critical moment is when the presenter makes a selection on a menu. From her point of view there is no need for the menu anymore and it is taken away. The system, however, can capture the selection event and pause the timeline so that the menu lingers on the public screen for an extended period suitable for passive viewers (possibly with animated highlighting of the selected menu item). It is also possible to identify the creation of a context menu, and smoothly move its clone to a neutral placement that does not block the operation's context (Figure 9a).

Watching interactions with menus is not the best way to convey operations to viewers. In many cases, replacing the menu selection with a different feedback, such as specifying the selection in a subtitle is better (Figure 9b). Consider a presenter who scrolls through a menu until finding a specific item. It is hard for a viewer to tell if the menu was closed because a selection was made or the menu was released with the ESC key. On the other hand, the subtitle scheme reports only when a selection is actually made. This scheme still works if the presenter uses keyboard shortcuts or gestural menus because semantic glue translates these back to a menu item description for display.

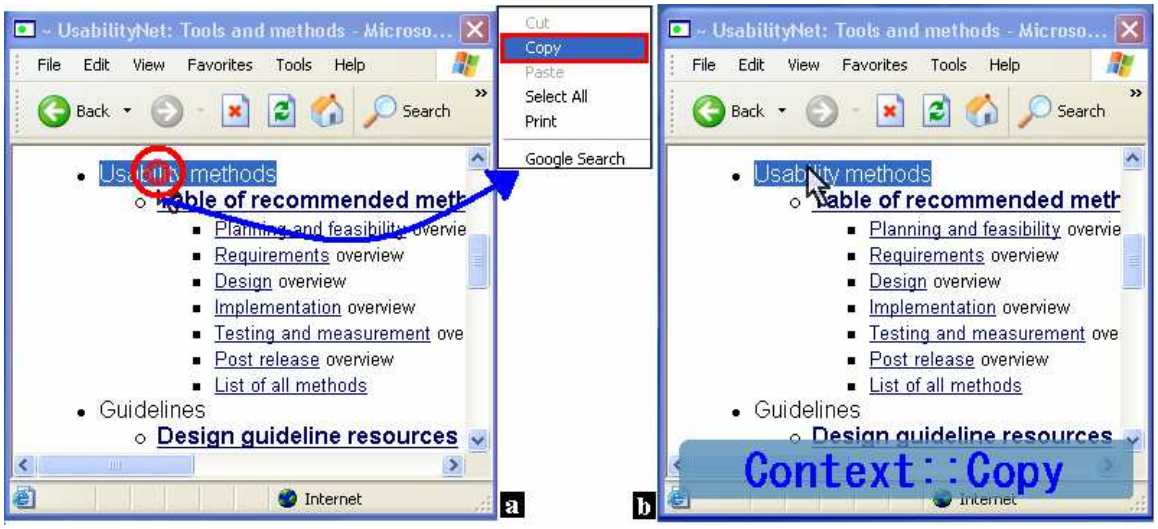


*Figure 9: Manipulating menus: (a) moving a context menu to the side + highlighting selection, (b) replacing a menu selection with a subtitle.*

## Mouse Cursor manipulations

We are able to provide viewers with a different representation for the mouse cursor to better suit their needs, while the presenter can continue working with the normal cursor.

Some of these alternative representations are shown in Figure 10. They assist viewers in tracking the cursor and the presenter's focus of interaction. Alternative cursors can support augmented or relaxed verbosity. They can increase verbosity by visually encoding low-level interactions (such as mouse button clicks or keyboard presses) or reduce it by displaying only the state indicators discussed previously. Finally, it may be desired to conceal the mouse cursor altogether when the presenter interacts with components that are private so mouse positions do not disclose interactions.

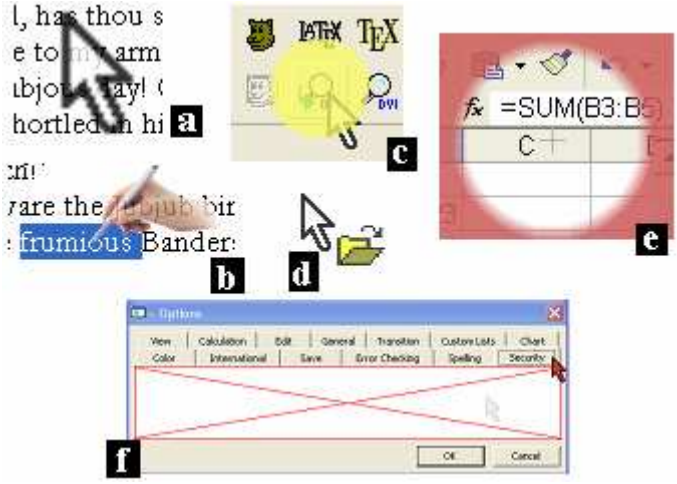


*Figure 10: (a and b) changing cursor size, shape and transparency, (c) visual feedback on the presenter's mouse clicks, (d) a private state indicator embedded on the cursor, (e) spotlight focus+context cursor, and (f) the cursor is hidden when going into the privacy-protected security tab.*

### Access Control

Tools like [11], [7] and remote desktop solutions enable the presenter to give a viewer control over keyboard and mouse input to her machine (coarse floor control). As far as the OS is concerned the input is coming from the presenter and therefore a viewer can do whatever the presenter can do. This is highly undesirable, since a viewer may make changes to non-shared applications or make unwanted changes to the application shared. We integrated the basic functionality of PointRight into our system, yet we are able to use the semantic layer to identify locations of widgets, menus and controls that should not be accessed. Thus when a viewer sends a mouse click event on such a control, the system will not pass this event to the OS. Similar treatment can be applied to keyboard events (this requires more queries on the application). In this way a fine grained access control policy can be applied to applications that were not designed for it.

## *Related Work*

There is a large corpus of research on collaboration-aware tools and frameworks that allow custom view sharing, including commercial tools (see a survey in [3] or [16]). Our system is targeted at existing collaboration transparent single-user tools, that were not designed for multiple views and cannot accommodate code changes. In this context, it is useful to classify these based on a mapping of the sharing architecture space. The Zipper model, presented in [4], looks at the common layers comprising an application: User, Screen, Window, Widget, View, Model to determines the share branching level. In each architecture one of these layers is the branching point. All layers below it are shared and all layers above it are replicated.

Closest to our approach are tools that replicate the screen or window layers (known as centralized tools), such as VNC [11], NetMeeting[4] or XTV [1] (replicating X-Windows display commands). VNC forces sharing of the entire screen, including windows of applications a presenter might like to keep private. NetMeeting and XTV share all windows of a specific application, but lack the ability to keep some dialogs or palettes private. In terms of supporting different views, NetMeeting allows a presenter to "pause and play" the view sharing, which could be used for concealing private states or interactions. However, this is manually controlled and very error-prone, leaving the presenter with the onus to identify privacy concerns, some of which are unpredictable. All of these tools lack any ability to change the contents of the replicated screen parts and therefore cannot handle sub-window elements or provide highlighting.

Flexible JAMM [3] is a framework for transparently replicating the widget layer. It is based on replacing some widgets and components of existing single-user applications with multi-user versions that can be synchronized with some degree of differentiation. This approach puts different constraints on the running environment and underlying code that make it unsuitable for commercial off-the-shelf applications.

[4] http://www.microsoft.com/windows/netmeeting/

Recently some research efforts have focused on synchronizing two running single-user applications (i.e. replicating the Model layer), using operational transformation techniques. A prominent example, working with an off-the-shelf editor (MS Word) is CoWord [16]. Each user has independent control of her copy while the system synchronizes the underlying document models using the application's API (typically without support for private document parts). This solution assumes complete independence of views, which is not suitable for presentation scenarios. Still, with some effort views can be synchronized as well (replicating the View layer). The solution requires both presenter and viewer to have a copy of the application (or to run two copies of the application on a single computer, which is not always possible).

The major drawback of JAMM, CoWord and other replication-based solutions is the need to synchronize the different application replications. Synchronization can be a very hard Problem [4], requiring resource locking or forcing expensive calculations to be carried out multiple times.

Wincuts [15] is a bitmap-based system that provides some spatial manipulation of shared windows. It allows a presenter to manually select a region of the window and publish only that part. Similar ideas are presented in regards to window layouts and screen space use in [6]. While still allowing a lot of flexibility and addressing clutter and privacy issues, this approach quickly breaks down when a presenter needs to resize or scroll a window. Our system completely subsumes these approaches, automating spatial manipulations and blending them with other filters.

A limited set of manipulations to the application's visual surface were presented in [9]. Some of the manipulations are similar to ours (e.g. region highlighting), but they are entirely targeted at supporting the work of a single user. Our system demonstrates how these can be used in a multi-user shared view scenario. On the technical side, this approach requires overriding some of the low-level drawing routines, consistent ordering and grouping of component drawings and cannot reason about information that does not go through the display pipeline (like field names). These demands are not fully met by off-the-shelf tools and are very hard to support under some operating system (Windows). Still, this may be a viable way for performing some of the semantic glue operations and other manipulations. Our plug-in architecture allows us to incorporate these techniques into our system.

In regards to privacy and clutter, Pebbles [8] offers to replicate some application components on a handheld device. Thus a presenter may choose to conduct some interactions on her handheld or auxiliary computer to avoid exposure. This approach requires extra hardware to be present and it does not address viewer needs for awareness cues. Most importantly this approach is limited in the type and complexity of components that can be recreated on the handheld (text fields and menu work well, but part of a worksheet relying on other spreadsheet parts may not).

Research into single display privacyware has resulted in several platforms enabling each user to have a different view of a shared display ([12]). However, these systems still

require running software that is capable of supporting differentiated views, so collaboration-aware tools must be used. Our solution enables using such systems with off-the-shelf applications.

Finally, commercial presentation authoring and playback tools such as Microsoft PowerPoint or Apple Keynote have recently taken advantage of multi-display technology to play the presentation on a public screen, while providing a private view to the presenter (where she can view her notes or check other slides). Our approach can provide similar advantages to other single-user applications.

## *Future Work and Discussion*

We expect to improve the performance of our system by relying on window paint events for updates rather than on a timer, and to support distributed view sharing based on a variant of the RFB protocol. When considering security, access control and privacy protection as well as performance, we intend to look at different models of where blurring and other window manipulations occur (presenter's machine, viewer's machine, a trusted third party server).

We need a higher degree of automation in extracting private elements, states, and hints on what level of verbosity to use. One possible direction is applying user modelling and machine learning techniques to learn what elements are considered private by a presenter and apply these to search the widget space and document model. Another possible direction is harnessing ``programming by example'' techniques.

We will run a formal user study on the next version of the system to evaluate its benefit to viewers and presenters. To date we have verified the problems identified with informal studies and tested the framework with several collaboration practitioners.

## *Conclusions*

We have introduced a unified solution for privacy concerns and verbosity control to assist a presenter and her audience in generalized presentation scenarios. These growing concerns are not addressed by current single-user application sharing modes. Our design introduces role-driven views for each type of participant, balancing between the presenter's privacy needs and the audience's awareness needs. Our system is based on applying image filters, spatial and timeline manipulations to bitmap representations of shared windows in contrast to similar techniques that have been recently shown to be unsuitable for addressing privacy in video [5].

The system's framework is general and works with off-the-shelf applications, requiring only a limited glue layer. A prototype of the system was created and tested with several commercial applications. As part of our work we discovered how Accessibility APIs can be leveraged to address privacy concerns. Our system allows additional intermediate

sharing layers to be introduced through an extensile plug-in architecture beyond the conventional screen, application or window layers of the Zipper model [4].

The system improves the quality of generalized presentation sessions. It protects a presenter from exposing private information and elements, allowing her to work normally and comfortably. It assists viewers in maintaining a suitable level of awareness and in better understanding the presenter's intentions.

## *References*


1. Abdel-Wahab, H. M., and Feit, M., Xtv: A framework for sharing X Window clients in remote synchronous collaboration. In Proceedings, IEEE Conference on Communications Software: Communications for Distributed Applications & Systems. ACM Press, 1991.

2. Baudisch, P., Cutrell, E., and Robertson, G., High-Density Cursor: A visualization technique that helps
users keep track of fast-moving mouse cursors. In Proceedings of Interact, pages 236–243, 2003

3. Begole, J., Rosson, M. B., and Shaffer, C. A., Flexible collaboration transparency: Supporting worker independence in replicated application sharing systems. ACM Trans. Comput.-Hum. Interact., 6(2), pages 95-132, 1999.

4. Dewan, P., Architectures for collaborative applications. In Trends in Software: Computer Supported Cooperative Work, pages 169-193, 1999.

5. Greenberg, S., Neustaedter, C., and Boyle, M., Blur filtration fails to preserve privacy for home-based video conferencing. ACM Transactions on Computer-Human Interaction (TOCHI), 2005.

6. Hutchings D. R., Stasko, J., Revisiting display space management: Understanding current practice to inform next-generation design. In Proceedings of the 2004 conference on Graphics Interface, pages 127-134., 2004.

7. Johanson B., Hutchins G., Winograd T., and Stone M., Pointright: experience with flexible input redirection in interactive workspaces. In Proceedings of the 15th annual ACM symposium on User interface software and technology, pages 227- 234. ACM Press, 2002.

8. Myers B. A., Peck C. H., Nichols J., Kong D., andMiller, R., Interacting at a distance using semantic snarfing. In Proceedings of the 3rd international conference on Ubiquitous Computing, pages 305-314, 2001.

9. Olsen, D. R., Hudson, S. E., Verratti T., Heiner J.M., and PhelpsM., Implementing interface attachments based on surface representations. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 191-198. ACM Press, 1999.

10. Palen L., and Dourish P., Unpacking privacy for a networked world. In Proceedings of the conference on Human factors in computing systems, pages 129-136. ACM Press, 2003.

11. Richardson, T., Stafford-Fraser, Q., Wood, K.W., and Hopper, A., Virtual Network Computing IEEE Internet Computing. 2(1), pages 33–38. 1998.

12. Shoemaker G. B. D., and Inkpen K. M., Single display privacyware: augmenting public displays with private information. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 522-529. ACM Press, 2001.

13. Stefik, M., Bobrow, D. G., Foster, G., Lanning S., and Tatar, D., WYSIWIS revised: early experiences with multiuser interfaces. M. ACM Trans. Inf. Systems 5 (2), pages 147–167. 1987.

14. Tan, D. S., Czerwinski M., Information voyeurism: social impact of physically large displays on information privacy. In extended abstracts on Human factors in computing systems, pages 748-749. ACM Press, 2003.

15. Tan D. S., Meyers B., Czerwinski, M., Wincuts: manipulating arbitrary window regions for more effective use of screen space. In Extended abstracts of the 2004 conference on Human factors and computing systems, pages 1525-1528. ACM Press, 2004.

16. Xia S., Sun D., Sun C., Chen D., and Shen H., Leveraging single-user applications for multi-user collaboration: the coword approach. In Proceedings of the 2004 ACM conference on Computer supported cooperative work, pages 162-171. ACM Press, 2004.