МОДЕЛЬ ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЯ КОРПОРАТИВНОЙ СЕТИ ПЕРЕДАЧИ ДАННЫХ

М.М. Монахова, Д.А. Шерунтаев, И.С. Марков, Д.В. Мазурок (Владимир)

Одной из проблем, возникающих при разработке данного рода систем, является сложность при выделении поведенческих характеристик и признаков, которые формируются при работе пользователя на АРМ. Задача усложняется и тем, что процесс работы за компьютером может быть различным в зависимости от выполняемых задач. Эталонные профили различных пользователей должны существенно отличаться друг от друга, тогда как текущие профили одного человека должны быть схожи. Следовательно, для распознавания и идентификации профилей, а также установления соответствия их определенным пользователям и их сравнения необходимо отыскать наиболее информативные характеристики, которые будут отличаться от пользователя к пользователю.

Таким образом, разрабатываемая модель предлагает представление пользователя, в виде слепка его поведенческих характеристик, отражающих его особенности работы за АРМ. Необходимы эффективные комплексные показатели значения параметров пользовательской активности, не теряющие в своей массе индивидуальных особенностей каждого из пользователей. Одним из интересных методов анализа параметров является построение пиктограмм типа «лиц Чернова», предложенный американским математиком Германом Черновым в 1973 г. Он основан на особенностях восприятия и практически мгновенной оценки человеческим мозгом образа человеческого лица в целом, т. е. фактически комплексной оценки таких показателей как форма лица, носа, глаз, их расположение и т. д. При этом наблюдателем очень быстро и четко идентифицируются малейшие отклонения (асимметрия элементов, их частичное увеличение и т. п.). По оценке ряда ведущих специалистов, к разряду важнейших характеристик лица можно отнести, например, глаза, губы, брови и др.

В результате исследований были определены следующие признаки, которые были использованы в рамках данной работы, которые, по мнению автора наиболее информативно характеризуют действия пользователя на компьютере:

1. Логин пользователя
2. Рабочее время пользователя
3. Средняя скорость движения мыши
4. Средняя координата движения мыши по оси X
5. Средняя координата движения мыши по оси Y
6. Среднее время между нажатием клавиши мыши и началом движения курсора
7. Среднее отклонение от траектории кратчайшего пути
8. Максимальное отклонение от траектории кратчайшего пути
9. Время нажатия буквенных клавиш
10. Скорость набора текста
11. История посещения сайтов
12. Запущенные приложения и службы

Отправной точкой для привязки характеристик к определенному пользователю является ничто иное как личный идентификатор сотрудника – Логин, используемый для соотношения набора поведенческих признаков с определенным человеком. Параметр – рабочее время, отображает промежуток времени, в которое данный пользователь должен находится за своим рабочим местом и выполнять положенные ему должностной инструкцией задачи. Нахождение на рабочем месте сверх установленных трудовым договором значений, без согласования с руководством, само по себе уже является нарушением и может быть возможным показателем инцидента информационной безопасности.

Исследование клавиатурного почерка

Основными показателями, характеризующими клавиатурный подчерк пользователя, являются скорость набора текста - Vtcp и время нажатия буквенных клавиш - delay, которые являются стабильными показателями, зависящими как от моторики человека, так и его навыков владения клавиатурой.

Для оценки информативности показателя delay в рамках данной работы было проведено несколько опытов для различных пользователей (таблица 1).

Таблица 1 – Время удержания клавиш

Пользователи	Время удержания клавиш - **delay**				Среднее
	А	Б	…	Я	
User 1	0.07977867	0.0811137		0.0771137	0,0799
User 2	0.08986563	0.0911001		0.0889881	0,0899
User 3	0.02094322	0.0300012		0.0192341	0,0233

Как можем заметить исходя из таблицы, временные различия в нажатии буквенных клавиш в рамках одного пользователя колеблются незначительно, таким образом предложение использовать в качестве поведенческих характеристик пользователя отдельно взятые клавиши не приведет к нужному результату, однако усредненное значение по всем буквенным клавишам значительно отличается от пользователя к пользователю (User 1 – 0.0799, User 2- 0.0233), что позволяет использовать данную характеристику для создания модели представления пользователя. Подсчет данной величины совершается через определенные интервалы времени и рассчитывается по формуле:

$$\boldsymbol{delay}=\Sigma\boldsymbol{delayii}; \quad (\mathbf{1})$$

где **i** – количество регистраций параметра.

Параметр скорость набора текста - Vtcp отображает уровень владения пользователя клавиатурным вводом. Навыки скоро печати значительно отличаются от человека к человеку. Скорость ввода текста пользователя, владеющего техникой «слепой печати» будет значительно выше по сравнению с пользователем, наблюдающим в процессе набора текста за движением пальцев рук. Так же, различия могут быть обусловлены биологическими факторами: различными заболеваниями суставов, возрастом пользователя и прочими, что так же позволяет использовать данный признак в качестве информативного для построения модели (Рисунок 1)

Регистрация данного параметра может вестись исключительно в моменты работы с текстовым окном, в противном случае система может принять в качестве результатов скоро печати нажатие функциональных клавиш и клавиш быстрого доступа (Формула 2).

$$\boldsymbol{Vt}\text{ср}=\Sigma\boldsymbol{Vt}\text{ср}\boldsymbol{ii} \quad (\mathbf{2})$$

где **i -** количество регистраций параметра в момент работы с текстом.

#	Имя	Дата время	Набрано слов/ знаков	Длительность сек.	Точность %	Слов в мин.	Знаков в мин.
1	Студент	04.02.2019 14:13:41	28/260	92	17	18	170
2	Студент	04.02.2019 14:13:32	28/260	83	17	20	189
3	Студент	04.02.2019 14:13:20	28/260	71	17	24	221
4	Студент	04.02.2019 14:13:09	28/260	60	17	28	262
5	Студент	04.02.2019 14:13:03	57/465	131	100	26	214
6	Студент	04.02.2019 14:12:38	28/260	29	17	58	542
7	Студент	04.02.2019 14:12:29	28/260	66	59	26	237
8	Enferon	04.02.2019 14:11:45	57/465	52	70	66	541
9	Студент	04.02.2019 14:11:38	29/282	45	100	38	372
10	Студент	04.02.2019 14:10:16	30/264	45	22	40	353

Рис. 1 – Результаты тестирования клавиатурного почерка среди студентов

История посещения сайтов

Для формирования модели профиля поведенческих характеристик пользователя, так де предлагается использование таких показателей как история посещенных сайтов **H** и запущенные приложения / службы **S.** Данные признаки за короткий промежуток времени являются слабо информативными, непостоянство использования как ресурсов интернета, так и запуска вспомогательного ПО не позволяет однозначно соотнести их с пользователями, однако, как показали опыты, при увеличении времени сбора данных по этим параметрам проявляется закономерность в использовании тех или иных веб-ресурсов, а так же запуске программного обеспечения.

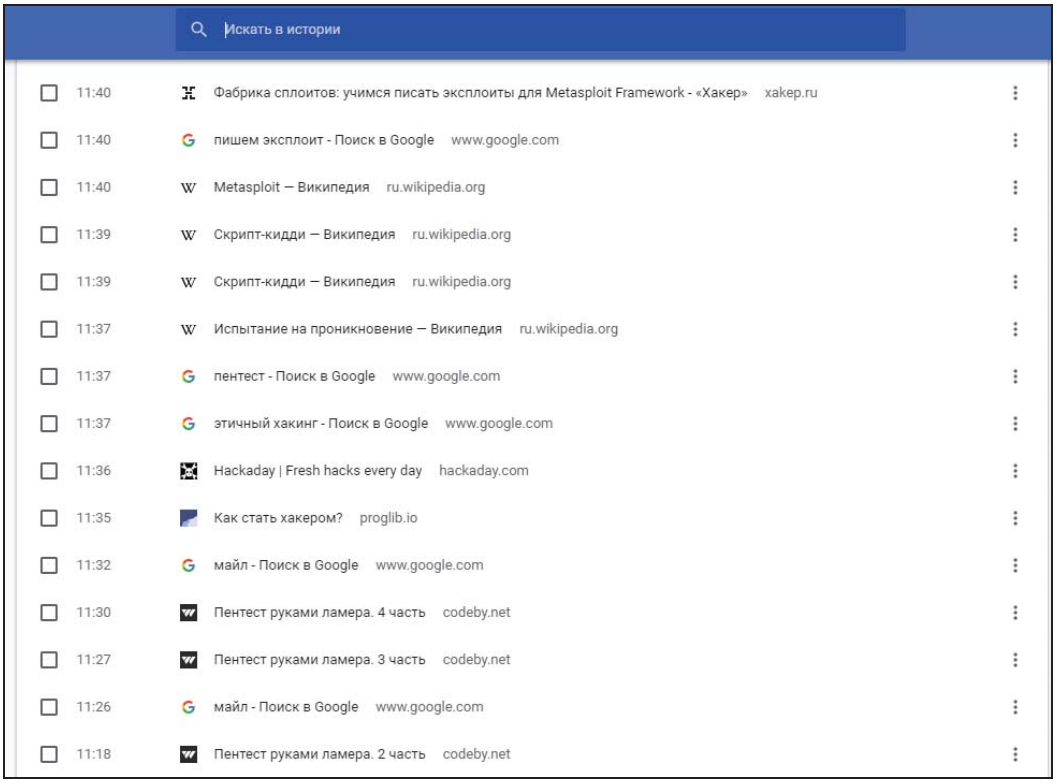


Рис. 2 – История посещения сайтов

Используемые программы

Пользователи в большинстве своем пользуются теми же инструментами, и посещают те же интернет страницы в своей работе, которые посещали ранее. Таким образом, данные признаки становятся информативными, при достаточно длительном времени сбора данных эталонного профиля.

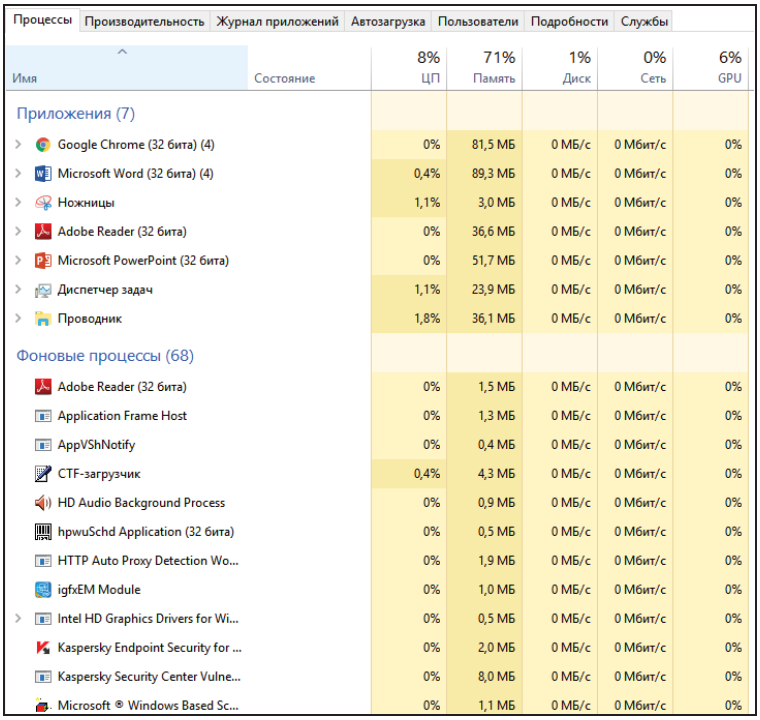


Рис. 3 – основные используемые процессы и программы

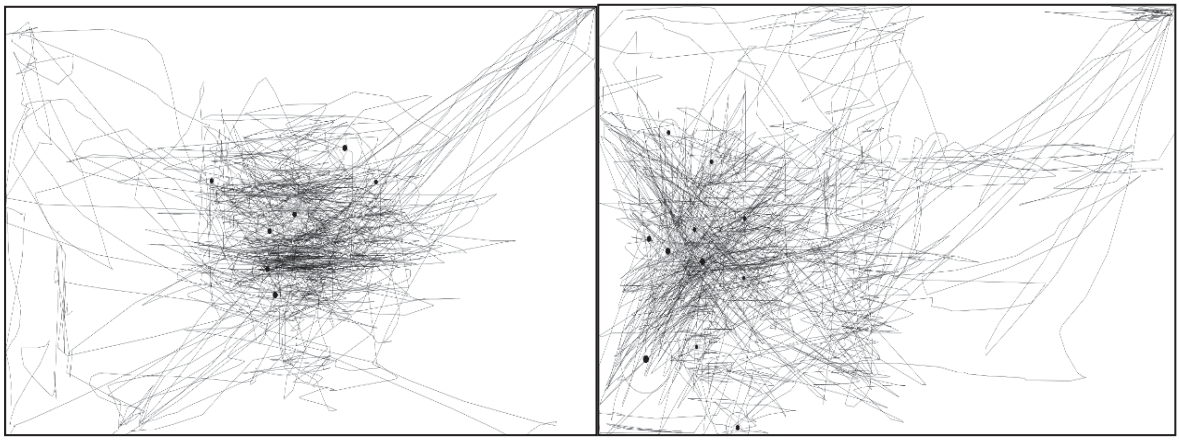
Принципиальное различие данных параметров заключается лишь в том, что запуск одного лишнего приложения пользователем может рассматриваться как инцидент информационной безопасности, тогда как переход на сайт, не входящий в эталонный профиль, не будет являться нарушением.

Работа с мышью

В качестве признаков, регистрируемых при работе с мышью, были использованы следующие: Средняя координата движения мыши по осям X и Y, средняя скорость движения мыши, время между нажатием ЛКМ и началом движения курсора, среднее и максимальное отклонение от траектории кратчайшего пути.

Средние координаты движения мыши $\boldsymbol{X}$ср ,$\boldsymbol{Y}$ср характеризует типовую активность пользователя в течении рабочего дня, области просмотра у различных сотрудников в зависимости от рода деятельности существенно отличаются, так если бухгалтер в течении рабочего дня в основном занимается печатью и правкой документов профиль его перемещений мыши будет близок к центральной области экрана, тогда как у начальника связи – размыт по краям, ввиду специфики интерфейса отображения сайтов и оболочки операционных систем.

На Рисунке 4 (а, б) представлены тестовые профили движения курсора пользователей А и Б за один рабочий день.



(а) (б)

Рис. 4 – Профили движения

Пользователь А в течении всего рабочего дня пользовался редактором документов Microsoft Word, учитывая специфику данного ПО график движения курсора мыши сосредоточен в центральной части рисунка и практически отсутствует по верхнему и правому краю, так как именно в центральной части сосредоточена рабочая область программы. Напротив график пользователя В сосредоточен в левой области, данный пользователь в течение тестового периода пользовался редактором таблиц Microsoft Excel. Расчет данных параметров производится по формулам:

$\boldsymbol{X}$ср=Σ$\boldsymbol{xiii}$; (**3**)

$\boldsymbol{Y}$ср=Σ$\boldsymbol{yiii}$; (**4**)

где **i** - количество регистраций параметра за отведенный промежуток времени

Несомненно, расчет данного параметра необходимо проводить только в переделах активного окна запущенного приложения, в противном случае усредненные значения будут стремится к центральной точке монитора автоматизированной рабочей станции.

Скорость движения курсора мыши $\boldsymbol{Vm}$ср, между элементами интерфейса вычисляется по формуле:

$\boldsymbol{Vm}$ср=(Σ$\boldsymbol{tend}-\boldsymbol{t1di}$)$\boldsymbol{i}$; (**5**)

где: $\boldsymbol{t1}$- время начала движения курсора; $\boldsymbol{tend}$– время окончания движения курсора; $\boldsymbol{d}$– расстояние между элементами интерфейса; $\boldsymbol{i}$- количество регистраций параметра за определенный временной интервал.

Изначально в данной формуле присутствовал еще один параметр – $\boldsymbol{t0}$, фиксирующий время нажатия левой клавиши мыши на функциональный элемент интерфейса, однако данный параметр правильнее использовать в рамках отдельного признака, смысл которого - время между нажатием ЛКМ и началом движения курсора, который рассчитывается по формуле:

$\boldsymbol{T}$срΣ$\boldsymbol{t1}-\boldsymbol{t0ii}$; (**6**)

где **i** - количество регистраций параметра за отведенный промежуток времени.

В качестве элементов, между которыми производится измерение параметров используются функциональные элементы активных окон программного обеспечения, а также функциональные элементы управления Windows, между которыми происходит перемещение курсора. Иллюстрированное изображение данных параметров приведено на рисунке 5

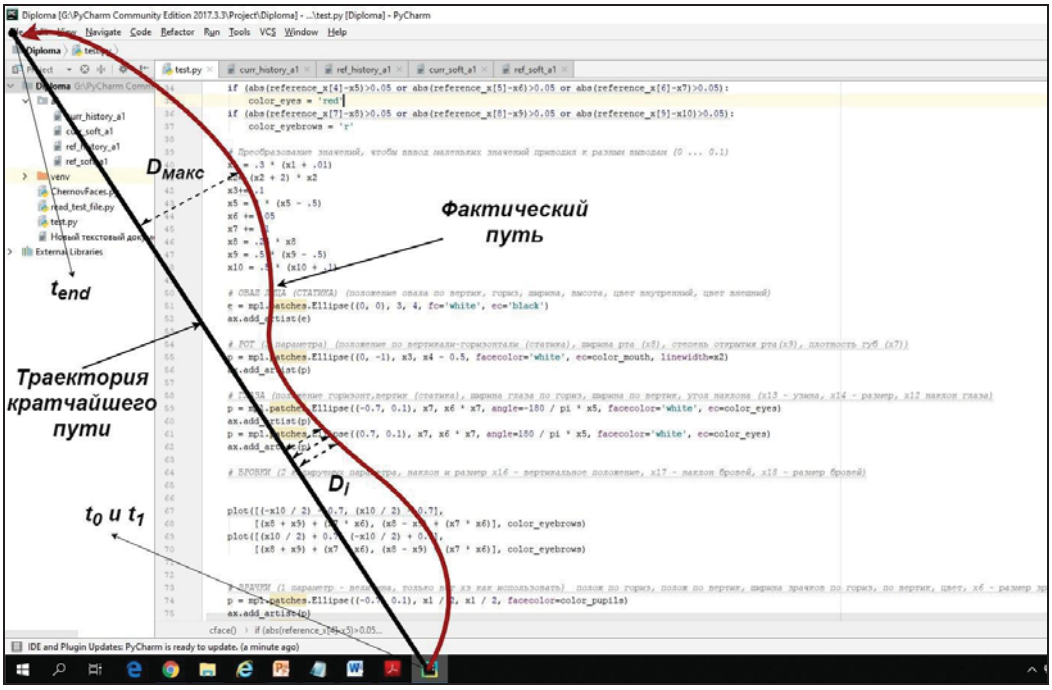


Рис. 5 – траектория кратчайшего пути

В момент нажатия по функциональному элементу Windows (в момент времени $t0$) запускается временной отсчет, который отменяется спустя секунду, если движение не началось (т.к. секундное ожидание может свидетельствовать о простое курсора, либо вводе данных в текстовую форму). В ином случае временной промежуток до начала движения (до момента $t1$) становится равным разнице временных величин $t1-t0$. Затем осуществляется движение курсора к целевому элементу интерфейса и в момент времени $tend$, при повторном нажатии левой кнопки мыши, отсчет завершается. В результате чего получаем одну регистрацию параметра, усредняя полученные значения за временной интервал получаем среднее значение поведенческого параметра Vmср.

Вычисление отклонений Dср и $Dмакс$ от траектории кратчайшего пути так же представлено на рисунке выше… и вычисляется по формулам

$$D\text{ср}=\Sigma Diii; \quad (7)$$

$$D\text{макс}=\Sigma D\text{макс}iii; \quad (8)$$

где **i** - количество регистраций параметра за отведенный промежуток времени.

Данные параметры демонстрируют уровень владения пользователя мышью, его координацию. Исследования показали, что Dср и Dмакс значительно отличаются от пользователя к пользователю, что позволяет их использовать в качестве поведенческих характеристик в разрабатываемой модели.

Формирование эталонного профиля легитимных пользователей будет осуществляться в процессе работы за автоматизированным рабочим местом. Усреднение данных по эталонным значениям будет проводится за промежуток времени равный одной рабочей неделе (5 дней). С 8 утра до 18 вечера. Данные по представленным формулам, для текущего мониторинга будут усредняться за один час работы системы и сравниваться с параметрами эталонного профиля. Для визуализации и удобства анализа, а также быстрого реагирования на изменение параметров, поведенческие характеристики будут визуализироваться графиками виде «Лиц Чернова» отображающих изменения параметров увеличивающимися или уменьшающимися размерами элементов лица.