Πανεπιστημιο Πειραιως
Τμημα Ψηφιακων Συστηματων

Εργαλεία και Μέθοδοι Συλλογής και Πιστοποίησης Περιεχομένου Ιστοσελίδων

Μεταπτυχιακή Διπλωματική Εργασία στα πλαίσια του Μ.Π.Σ.:
«Τεχνοοικονομική Διοίκηση και Ασφάλεια Ψηφιακών Συστημάτων»

Ευθυμιος Γ. Λαλας
(α.μ. μτε1113)

Τριμελής Επιτροπή:

Σ. Κατσικας	Κ. Λαμπρινουδακης	Χ. Ξενακης
Καθηγητης	Αναπλ. Καθηγητης	Επικ. Καθηγητης
	(Επιβλεπων)	

Πειραιάς, Σεπτέμβριος 2013

Περίληψη

Είναι αναμφησβήτητο γεγονός ότι στις μέρες μας εγκλήματα διαφόρων ειδών διαπράττονται με τη χρήση ή εναντίον ηλεκτρονικών συσκευών, δικτύων ή άλλων μέσων επικοινωνίας. Για να αντιμετωπιστεί η νέα αυτή κατάσταση, τα σύγχρονα εργαλεία εγκληματολογικής ανάλυσης (forensic tools) έχουν εξελιχθεί, ενσωματώνοντας δυνατότητες χειρισμού σχεδόν όλων των ειδών ψηφιακού περιεχομένου. Παρ΄ όλα αυτά, η συλλογή και η πιστοποίηση περιεχομένου ιστοσελίδων για αποδεικτικούς λόγους παραμένει ανοιχτό πρόβλημα, κυρίως λόγω της αδόμητης και δυναμικής φύσης του διαδικτύου.

Η συνηθισμένη πρακτική για τη συλλογή περιεχομένου ιστοσελίδων είναι η αποθήκευση ("Save As") ή αποτύπωση ("Screenshot") του περιεχομένου από τον ενδιαφερόμενο ή από κάποιον τρίτο, ο οποίος μετά υποστηρίζει στο δικαστήριο το περιεχόμενο που καταθέτει. Άλλες μέθοδοι συμπεριλαμβάνουν την μαγνητοσκόπηση της επίσκεψης σε μία ιστοσελίδα, την χρησιμοποίηση τρίτων (ελεύθερων ή εμπορικών) υπηρεσιών από το διαδίκτυο, καθώς και τη λήψη κατάθεσης από τον δημιουργό της ιστοσελίδας.

Στη παρούσα εργασία, αναλύουμε τις νομικές πτυχές της χρησιμοποίησης περιεχομένου ιστοσελίδων ως αποδεικτικό μέσο και καταγράφουμε όλα τα κριτήρια για την ασφαλή και χωρίς αμφισβήτηση αποθήκευση αυτού. Στη συνέχεια εξετάζουμε κατά πόσο και σε ποιο βαθμό οι υπάρχουσες μέθοδοι πληρούν τα προαναφερθέντα κριτήρια. Τέλος παρουσιάζουμε μία νέα μέθοδο, ονομαζόμενη ProCAVE, που αποδεικνύεται ότι πληρεί όλες τις προϋποθέσεις για την - πέρα από κάθε αμφισβήτηση - αποθήκευση και πιστοποίηση περιεχομένου ιστοσελίδων.

Γίνεται μνεία ότι μέρος της παρούσας εργασίας, ως επιστημονικό άρθρο έγινε αποδεκτό για δημοσίευση στα πρακτικά του 10th International Conference on Trust, Privacy & Security in Digital Business (TrustBus 2013) [17].

Abstract

It is an undisputable fact that nowadays many different types of crime are conducted by utilizing some type of electronic device - communication. To address this new situation, modern forensics tools evolved, becoming sophisticated enough to handle almost all kinds of digital content. However, collecting and validating the authenticity of online content remains, until now, a problem to resolve. This has to do with many reasons, but mainly with the dynamic and unstructured nature of the internet and the web.

The common practice for preserving online content for evidentiary purposes is to capture (screen-shot) or save a web page, the authenticity of which is usually validated in a judicial process by an expert's testimony. Other methods suggest video recording a browsing session, utilizing free or commercial services for content preservation, or having the creator of the web page testify about its content at a specified date and time.

In this thesis, we analyze the legal aspects of presenting online content as evidence, and we list all criteria for acceptable online content preservation. Next, we present all available methods for content preservation and we see how they fit to the criteria previously stated. Finally, we introduce PRoCAVE, a simple software architecture with a set of accompanying procedures, and we argue that their combined use can deliver evidence from online sources in the court, in a sound and privacy-preserving manner.

Part of this work was accepted for publication in the proceedings of the 10th International Conference on Trust, Privacy & Security in Digital Business (TrustBus 2013) [17].

Ευχαριστίες

Φτάνοντας στο τέλος των μεταπτυχιακών μου σπουδών, αισθάνομαι την ανάγκη να αναφερθώ στους ανθρώπους που, ο καθένας με το δικό του τρόπο, συνέβαλαν στην ολοκλήρωση αυτών, και να τους ευχαριστήστω.

Αρχικά, θα ήθελα να εκφράσω τις ευχαριστίες μου στον Αναπλ. Καθηγητή κ. Λαμπρινουδάκη Κωνσταντίνο για τη καθοδήγησή και τις χρήσιμες συμβουλές του σε όλη τη διάρκεια της μεταπτυχιακής εργασίας μου, αλλά και γενικότερα του μεταπτυχιακού προγράμματος. Οι συζητήσεις μαζί του και το ενδιαφέρον που επέδειξε υπήρξαν καταλυτικά, και για τον λόγο αυτό τον ευγνωμονώ.

Επίσης, ευχαριστώ θερμά τα λοιπά μέλη της τριμελούς επιτροπής, Καθηγητή κ. Κάτσικα Σωκράτη και Επίκουρο Καθηγητή κ. Ξενάκη Χρήστο για τη διδασκαλία τους στα πλαίσια του μεταπτυχιακού προγράμματος, τις γνώσεις που μου μετέδωσαν αλλά και την πολύτιμη βοήθειά τους. Στο ίδιο πλαίσιο, ευχαριστώ ιδιατέρως την Αναπλ. Καθηγήτρια κα Μήτρου Λίλιαν, για τη συνεργασία της στο ερευνητικό κομμάτι της παρούσας εργασίας, και τελικά στη διαμόρφωσή του σε δημοσιεύσιμο επιστημονικό άρθρο.

Τέλος, θερμές ευχαριστίες οφείλω στο Ίδρυμα Σταύρος Νιάρχος, για την χρηματοδότηση των διδάκτρων του μεταπτυχιακού προγράμματος, δίχως την οποία η συμμετοχή μου σε αυτό θα ήταν αδύνατη.-

Αθήνα, Σεπτέμβριος 2013
Ευθύμιος Λαλάς

Στη Β.
"Trust I seek and I find in you"

Περιεχόμενα

Κατάλογος Σχημάτων

Κεφάλαιο 1

Εισαγωγή

Στο εισαγωγικό αυτό κεφάλαιο θα δώσουμε κάποιους βασικούς ορισμούς για το ηλεκτρονικό έγκλημα, και θα αναφερθούμε στις προκλήσεις σχετικά με τη συλλογή και πιστοποίηση περιεχομένου ιστοσελίδων. Κατόπιν θα αναφερθούμε περιληπτικά στην λύση που προτείνουμε για την αντιμετώπιση του προβλήματος, και θα παραθέσουμε τη δομή της υπόλοιπης εργασίας.

1.1 Γενικά στοιχεία

Το διαδίκτυο σήμερα χρησιμοποιείται από δισεκατομύρια χρήστες σε όλον τον κόσμο, για επικοινωνία, διασκέδαση, ανταλλαγή πληροφοριών και για αγοραπωλησία αγαθών και υπηρεσιών. Παρ΄ όλα αυτά, αυτού του είδους η νέα πραγματικότητα έχει και μία «σκοτεινή πλευρά» [6]· πλέον, ένας μεγάλος αριθμός εγκλημάτων διαπράττονται διαμέσου ή με τη χρήση αυτού. Τα εγκλήματα αυτά συνήθως χωρίζονται στις τρεις παρακάτω κατηγορίες [4, 6]:

- **Εγκλήματα εναντίον υπολογιστικών συστημάτων (computer crimes):** είναι αυτά που έχουν σαν κύριο στόχο υπολογιστές, δεδομένα, ή συστήματα επικοινωνιών, και δεν θα υφίσταντο σαν αδικήματα αν δεν υπήρχαν οι υπολογιστές και το διαδίκτυο. Τέτοια είναι η παραβίαση υπολογιστικών συστημάτων, η άρνηση υπηρεσιών (denial of service) κλπ.
- **Εγκήματα σχετικά με υπολογιστικά συστήματα (computer-related crimes):** είναι τα κλασικά εγκλήματα που υπήρχαν ανέκαθεν, και απλώς εκτελούνται πλέον και μέσω υπολογιστικών συστημάτων. Τέτοια είναι η απάτη, η κλοπή πνευματικής ιδιοκτησίας κλπ.

- **Εγκλήματα περιεχομένου (content-related crimes):** ομοίως αφορούν κλασικά εγκλήματα, όμως στη συγκεκριμένη περίπτωση το αδίκημα σχετίζεται με κάποιο παράνομο περιεχόμενο, παρά με την ίδια την πράξη, όπως στην προηγούμενη περίπτωση. Τέτοια είναι η κατοχή και διακίνηση υλικού πορνογραφίας ανηλίκων, η δυσφήμιση μέσω διαδικτύου, κλπ.

Στο νομικό σύστημα των περισσότερων χωρών, η απόδειξη της ενοχής κάποιου ενώπιον κάποιου δικαστηρίου συνήθως αποτελεί συνάρτηση και συνυπολογισμός *μέσων* (means), *κινήτρου* (motive) και *ευκαιρίας* (opportunity) [5]. Αυτό σημαίνει ότι ο κατηγορούμενος, για να αποδειχθεί ένοχος θα πρέπει να είχε τουλάχιστον κίνητρο, να είχε τα μέσα να διαπράξει το έγκλημα για το οποίο κατηγορείται, αλλά και να είχε και την ευκαιρία να το διαπράξει.

Παρ΄ όλα αυτά, η ύπαρξη και των τριών αυτών στοιχείων δεν είναι πάντα ικανή προκειμένου να καταδικαστεί κάποιος· συνήθως χρειάζονται αρκετά πειστήρια (evidence) που θα αποδεικνύουν ότι ο κατηγορούμενος, έχοντας το κίνητρο όντως χρησιμοποίησε τα μέσα που είχε στην κατοχή του όταν του δόθηκε η ευκαιρία (ένα παράδειγμα φαίνεται στο [5]).

Σε ότι αφορά την αποδοχή των πειστηρίων, είναι γεγονός ότι τα περισσότερα νομικά συστήματα είναι προσαρμοσμένα στους παραδοσιακούς τύπους εγκλημάτων. Έτσι, οποιαδήποτε πληροφορία προέρχεται από υπολογιστές ή διαδίκτυο αυτόματα πρέπει να πείσει για την ορθότητα και για τη σχετικότητά της προκειμένου να γίνει αποδεκτή. Δηλαδή πρέπει να έχει συλλεχθεί βάσει επίσημων και αδιαμφησβήτητων διαδικασιών [18], και να συνδέεται άμεσα με το υπό διερεύνηση αδίκημα [6]. Σε ότι αφορά το έγκλημα, με την παραδοσιακή ερμηνεία του, για να αποδειχτεί η ενοχή ενός κατηγορούμενου για ένα συγκεκριμένο αδίκημα θα πρέπει να αναγνωριστούν, συλλεχθούν και να παρουσιαστούν τα ψηφιακά πειστήρια με τέτοιο τρόπο ώστε να γίνουν αποδεκτά στην αποδεικτική διαδικασία [15].

Βάσει των παραπάνω, γίνεται άμεσα αντιληπτό ότι το διαδίκτυο αποτελεί πλέον, εκτός από σκηνή εγκλήματος (crime scene), και πρόσφορο έδαφος για τη συλλογή κύριων και δευτερευόντων ψηφιακών πειστηρίων. Τα ψηφιακά πειστήρια αυτά είτε σχετίζονται με το έγκλημα αυτό καθ΄ αυτό (όπως στα εγκλήματα περιεχομένου), είτε σχετιζονται με το θύτη ή το θύμα του εγκλήματος, και επομένως η ύπαρξη και χρήση τους μπορεί να βοηθήσει σημαντικά τα δικαστήρια στη λήψη αποφάσεων.

Παρ΄ όλα αυτά, λόγω της μη ιεραρχικής και αδόμητης φύσης του διαδικτύου, τα πειστήρια που ενυπάρχουν στο διαδίκτυο είναι εξ΄ ορισμού εφήμερα και ευμε-

τάβλητα, παρ' όλο που σε γενικές γραμμές, με την υιοθέτηση της Ευρωπαϊκής νομοθεσίας για την ψηφιακή υπογραφή (Ευρωπαϊκή Οδηγία 2000/31/EC) τα ψηφιακά έγγραφα και γενικά το ψηφιακό υλικό έχει αποκτήσει κύρος και νομική υπόσταση παρόμοια με το έντυπο υλικό.

Έτσι, σχετικά με τα εγκλήματα όπου συγκεκριμένο περιεχόμενο που βρίσκεται στο διαδίκτυο παίζει σημαντικό ρόλο, η κύρια πρόκληση είναι η σωστή και πέρα από κάθε αμφισβήτηση συλλογή, διατήρηση και παρουσίαση αυτού. Δηλαδή, προκειμένου να πετύχει τον στόχο του, ο οποίος είναι η υποβοήθηση της προσπάθειας του δικαστηρίου να διαμορφώσει και να εξάγει ένα συμπέρασμα, το περιεχόμενο που βρίσκεται στο διαδίκτυο θα πρέπει να συλλεχθεί και να πιστοποιηθεί κατάλληλα.

1.2 Ορισμός του προβλήματος

Στο σημείο αυτό, βάσει των προαναφερθέντων και προκειμένου να ορίσουμε το πρόβλημα, θα δώσουμε ένα σενάριο που αντιμετωπίζεται στην καθημερινότητα:

Σενάριο. *Σε μία ιστοσελίδα, έστω* `www.abc.com`*, βρίσκεται συγκεκριμένο περιεχόμενο το οποίο πρέπει να χρησιμοποιηθεί σε κάποια αποδεικτική διαδικασία. Για τον λόγο αυτό, πρέπει να αποθηκευτεί και να πιστοποιηθεί η γνησιότητά του, καθώς και ο χρόνος αποθήκευσης.*

Στο σενάριο αυτό, κάποιος ενδιαφερόμενος βρίσκει σε μία συγκεκριμένη ιστοσελίδα περιεχόμενο το οποίο είτε είναι από μόνο του παράνομο (εμπίπτει δηλαδή σε μία από τις κατηγορίες που αναφέρεται στην προηγούμενη ενότητα), είτε απλώς συνδέεται με τον θύτη ή το θύμα μίας παράνομης πράξης. Το περιεχόμενο αυτό αναφέρεται ως *ψηφιακό πειστήριο* (digital evidence) ή απλώς *πειστήριο* [4].

Η σωστή συλλογή και πιστοποίηση της γνησιότητας είναι βασικά στοιχεία της εγκληματολογικής ανάλυσης. Ως *συλλογή* αναφέρεται η λήψη είτε του πρωτότυπου πειστηρίου είτε ενός ακριβούς αντιγράφου αυτού, ενώ με τον όρο *πιστοποίηση της γνησιότητας* (authentication) ή απλώς *πιστοποίηση* θα αναφερόμαστε στις μεθόδους απόδειξης ότι το συλλεχθέν πειστήριο είναι όντως το αυθεντικό είτε αντίγραφο του αυθεντικού [14].

Ως αποδεικτικά μέσα, τα πειστήρια χρησιμοποιούνται στα νομικά συστήματα των περισσοτέρων χωρών τους τελευταίους δύο σχεδόν αιώνες. Για την εισαγωγή των πειστηρίων αυτών ενώπιον αποδεικτικών διαδικασιών, όλες οι χώρες έχουν ενσωματώσει κατάλληλους κανόνες και διαδικασίες. Για παράδειγμα στις

Ηνωμένες Πολιτείες, και ειδικά για την πιστοποίηση του περιεχομένου ιστοσελίδων, οι κανόνες αυτοί έχουν κωδικοποιηθεί [9, 19] και κατηγοριοποιηθεί ως ακολούθως:

- 901(b)(1): Πιστοποίηση περιεχομένο από μάρτυρα με ειδικές γνώσεις.
- 901(b)(3): Κατάθεση ειδικού σχετικά με το περιεχόμενο
- 901(b)(4): Ειδικά χαρακτηριστικά μία ιστοσελίδας που καθιστούν αυτοπόδεικτη την προέλευση του περιεχομένου.
- 901(b)(7): Πιστοποίηση περιεχομένου λόγω ύπαρξης σε κυβερνητικές ιστοσελίδες
- 901(b)(9): Σύστημα ή διαδικασία ικανή να παρέχει αξιόπιστα αποτελέσματα.
- 902(5): Επίσημες εκδόσεις.

Το σημαντικό σημείο σχετικά με την πιστοποίηση των πειστηρίων, και κατά συνέπεια την ενίσχυση του αποδεικτικού τους ρόλου είναι ότι δεν μπορούμε να βασιστούμε σε μία από τις παραπάνω μεθόδους, ή σε οποιαδήποτε μέθοδο συστήνεται από το νομικό σύστημα κάθε χώρας. Αντίθετα, η χρησιμότητα και η αποδεικτικότητα των πειστηρίων σε κάθε υπόθεση συνήθως αξιολογείται από το κάθε δικαστήριο [19]. Έτσι, καθίσταται σαφές ότι όσες περισσότερες διαδικασίες πιστοποίησης χρησιμοποιηθούν, τόσο πιθανότερο είναι το δικαστήριο να αποδεχτεί την γνησιότητα των πειστηρίων [11], και αυτός είναι και ο σημαντικότερος λόγος που όλες οι υπάρχουσες μέθοδοι συλλογής και πιστοποίησης αποτυγχάνουν να παρουσιάσουν αδιαμφησβήτητα πειστήρια, όπως θα αναλύσουμε και ευθύς αμέσως.

Πρώτα απ' όλα, η αποθήκευση και εκτύπωση μίας ιστοσελίδας παράγει ένα πειστήριο η αυθεντικότητα του οποίου πρέπει να υποστηριχθεί σε οποιαδήποτε αποδεικτική διαδικασία από αυτόν που προχώρησε στην συλλογή του [29], μία μέθοδος που πολλές φορές είναι αμφισβητήσιμη, ειδικά όταν ο χρήστης είναι ιδιοτελής, δηλαδή έχει συμφέρον από την ύπαρξη ή όχι συγκεκριμένου περιεχομένου. Επιπλέον, ο ακριβής χρόνος που το περιεχόμενο αυτό ήταν διαθέσιμο δεν μπορεί να αποδειχτεί, καθώς δεν έχουν εφαρμοστεί συγκεκριμένες μέθοδοι χρονοσήμανσης (όπως προτείνεται στο [13]).

Όπως αναφέραμε, υπάρχουν συγκεκριμένα εμπορικά εργαλεία ή υπηρεσιές ([2, 7, 34]) υπόσχονται ότι παρέχουν έναν αξιόπιστο μηχανισμό συλλογής περιεχόμενου ιστοσελίδων. Με τις μεθόδους αυτές, ικανοποιείται η απαίτηση της χρονοσήμανσης· παρ΄ όλα αυτά, ο ενδιαφερόμενος πρέπει για άλλη μια φορά να υποστηρίξει την ύπαρξη του περιεχομένου αυτού στο δικαστήριο. Το κόστος είναι ακόμη ένα σημαντικό μειονέκτημα, καθώς επίσης και το γεγονός ότι όταν κάποιος συναντά κάποιο παράνομο περιεχόμενο κατά την πλοήγησή του στο διαδίκτυο, δεν είναι πάντα εύκολο ή άμεσο να αγοράσει κάποιο ακριβό λογισμικό ή υπηρεσία προκειμένου να το αποθηκεύσει.

Η κατάθεση κάποιου ειδικού, που συνήθως είναι ανεξάρτητος από την υπόθεση (όπως συμβαίνει με τους διαχειριστές ιστοσελίδων διατήρησης περιεχομένου, σαν το [31]) είναι επίσης μία πιθανή επιλογη. Παρ΄ όλα αυτά, οι εταιρίες που αποθηκεύουν περιεχόμενο, όπως η προαναφερόμενη, δεν αποθηκεύουν δυναμικό ή προσωποποιημένο περιεχόμενο, ή άλλες φορές το περιεχόμενο έχει αφαιρεθεί πριν το αποθηκεύσει η εταιρία. Επιπλέον, η ανεύρεση ενός τέτοιου ειδικού για να καταθέσει είνα ιδιαίτερα δύσκολη υπόθεση, ακόμα και για τις Ηνωμένες Πολιτείες [3], πόσο μάλλον για χώρες στις άλλη άκρη του Ατλαντικού.

Η τελευταία λύση είναι να εξευρεθεί ο ιδιοκτήτης της ιστοσελίδας, και να καταθέσει σχετικά με το περιεχόμενο της το οποίο ήταν ανηρτημένο σε κάποια δεδομένη χρονική στιγμή. Αλλά αυτό γίνεται επίσης μία ιδιαίτερα δύσκολη διαδισία. Ακόμα όμως και να επιτρυχθεί, χρειάζονται για άλλη μια φορά εργαλεία και μέθοδοι για την τελική πιστοποίηση του περιεχομένου [16].

Είναι γνωστό σε όσους ασχολούνται με αποδεικτικές διαδικασίες ότι οι προσεγγίσεις που αναφέραμε παραπάνω προκαλούν πολλές συγχύσεις και αμφιβολίες που συχνά οδηγούν στην μη αποδοχή των πειστηρίων στο δικαστήριο, ή στην κατάρριψη της αποδεικτικής ισχύος τους. Ειδικότερα, κάποια επιχειρήματα που χρησιμοποιούνται συχνά για να αμφισβητηθούν τα πειστήρια είναι τα εξής: «ο μάρτυρας τροποποίησε την αποθηκευμένη ιστοσελίδα», «ο μάρτυρας άλλαξε τον HTML κώδικα της αποθηκευμένης ιστοσελίδας», «η ιστοσελίδα δεν ήταν οραατή εκείνη την ημερομηνία και ώρα», «ο ιδιοκτήτης της ιστοσελίδας είναι σε κάποια άλλη χώρα και δεν μπορεί να καταθέσει» κλπ.

Τέτοιοι ισχυρισμοί καταδεικνύουν ότι το κυρίαρχο στοιχείο έτσι ώστε τα πειστήρια να έχουν ισχυρή αποδεικτική αξία όπως προαναφέραμε είναι η πιστοποίησή τους, δηλαδή η χρήση εργαλείων και μεθόδων που θα αποδεικνύουν - πέρα από κάθε αμφισβήτηση - τον τρόπο με τον οποίο έχουν συλλεχθεί, αποθηκευθεί, χρονοσημανθεί και υπογραφεί τα πειστήρια. Μόλις συμβεί αυτό, όπως αναφέρεται και στο [8], τότε όλα τα υπόλοιπα προβλήματα αποδεικτικότητας είναι αυτά που οι δικηγόροι αντιμετωπίζουν συνεχώς στις αποδεικτικές

διαδικασίες· δηλαδή όχι πλέον αν τα πειστήρια είναι αυθεντικά, αλλά μόνο αν σχετίζονται με την υπόθεση, αν αποδεικνύουν το έγκλημα κ.α.

Μία επιπλέον λειτουργία που απουσιάζει από όλες τις τεχνικές και μεθόδους συλλογής και πιστοποίησης περιεχομένου, και εν μέρει μας έδωσε κίνητρο για την παρούσα εργασία, είναι η έννοια της *ιδιωτικότητας* (privacy). Πιο συγκεκριμένα, όπως αναφέρεται και στο [15], η χρήση εργαλείων συλλογής πειστηρίων και ψηφιακής εγκληματολογικής ανάλυσης μπορεί σε κάποιες περιπτώσεις να οδηγήσει σε παραβίαση του θεμελιώδους δικαιώματος της ιδιωτικότητος. Αυτό συμβαίνει διότι στα πειστήρια ενδέχεται να αποτυπώνονται ή να εμπεριέχονται προσωπικά στοιχεία του υπό διερεύνηση αντικειμένου, τα οποία αποκαλύπτονται στους συμμετέχοντες στη διαδικασία χωρίς να αφορούν απαραίτητα την υπόθεση.

Τέλος, όπως έχει αναφερθεί εκτενώς τα τελευταία χρόνια [10, 15, 21], τα εργαλεία εγκληματολογικής ανάλυσης σε γενικές γραμμές υστερούν σε τυποποίηση, με αποτέλεσμα να δημιουργούνται πολλά διαδικαστικά προβλήματα. Συγκεκριμένα για τη συλλογή και πιστοποίηση του περιεχομένου ιστοσελίδων, η τυποποίηση αυτή είναι πρακτικά αδύνατη με τις υπάρχουσες μεθόδους, καθότι αυτές είναι πολύ διαφορετικές μεταξύ τους, και εξαρτώνται ιδιαίτερα στις τεχνικές ικανότητες κάθε χρήστη.

1.3 Προσέγγιση

Έχοντας κατά νου την παρούσα κατάσταση, όπως την έχουμε περιγράψει έως τώρα, στην παρούσα εργασία θα προτείνουμε ένα καινοτομικό τρόπο συλλογής και πιστοποίησης περιεχομένου ιστοσελίδων, έτσι ώστε να αυτό να ικανοποιεί τους αποδεικτικούς σκοπούς συλλογής του. Η προσέγγισή μας, με την ονομασία ProCAVE (Privacy-preserving Collection and Authenticity Validation of online Evidence), στηρίζεται στην έννοια του web proxy [30], που δεν έχει χρησιμοποιηθεί ξανά στο παρελθόν για τον σκοπό αυτό.

Έτσι, υποστηρίζεται ότι μόλις ένας ενδιαφερόμενος αναγνωρίσει κάποιο περιεχόμενο στο διαδίκτυο το οποίο είτε είναι από μόνο του παράνομο, είτε σχετίζεται με κάποιο θύμα ή θύτη ενός εγκλήματος, τότε μπορεί να «πλοηγηθεί» στην ιστοσελίδα αυτή μέσω του web proxy. Ο ενδιάμεσος αυτός εξυπηρετητής θα επιστρέψει το περιεχόμενο στον ενδιαφερόμενο και ταυτόχρονα θα το αποθηκεύσει τοπικά, εφαρμόζοντας επιπλέον μεθόδους χρονοσήμανσης και ψηφιακής υπογραφής. Σε όποιο σημείο το επιθυμήσει, ο ενδιαφερόμενος μπορεί να ανακτήσει το περιεχόμενο από τον web proxy υπό τη μορφή πειστηρίου, έτοιμο

προς χρήση σε αποδεικτική διαδικασία.

Πλέον των ανωτέρω, με την προτεινόμενη λύση, η ίδια αίτηση HTTP μπορεί να πραγματοποιηθεί από πολλούς web proxies, έτσι ώστε το ίδιο περιεχόμενο να αποθηκεύεται ως πειστήριο ταυτόχρονα σε πολλά σημεία. Επίσης, όπως θα δείξουμε, στη λύση μας έχουμε ενσωματώσει και την έννοια της προστασίας της ιδιωτικότητας, έτσι ώστε ο ενδιαφερόμενος να μπορεί να αλλοιώσει νόμιμα το περιεχόμενο μιας ιστοσελίδας, έτσι ώστε να αποκρύψει στοιχεία η απόκαλυψη των οποίων θα συνιστούσε παραβίαση της ιδιωτικότητάς του. Φυσικά το αρχικό περιεχόμενο θα εξακολουθεί να υφίσταται αποθηκευμένο, για λόγους αναφοράς.

Για να αποδείξουμε τη χρησιμότητα και την καλή λειτουργικότητα της προτεινόμενης λύσης ProCAVE, προχωρήσαμε στην πιλοτική υλοποίησή της αναπτύσοντας ένα σύστημα με τις αρχές που περιγράψαμε. Τέλος, όπως θα δούμε, πραγματοποιήσαμε πειράματα χρήσης του συστήματος ProCAVE με μία σειρά από ιστοσελίδες, και τα αποτελέσματα απέδειξαν την ορθότητα των ισχυρισμών μας.

1.4 Δομή της εργασίας

Το υπόλοιπο της εργασίας έχει ως ακολούθως: στο Κεφάλαιο 2 θα κάνουμε μία αναφορά στις υπάρχουσες μεθόδους που έχουν επικρατήσει για τη συλλογή και πιστοποίηση του περιεχομένου ιστοσελίδων. Στο Κεφάλαιο 3 θα αναλύσουμε τη μέθοδο που προτείνουμε, επονομαζόμενη ProCAVE, και θα δείξουμε ότι υπερισχύει όλων των τωρινών μεθόδων. Στο Κεφάλαιο 4 θα αναφερθούμε σε μία πιλοτική υλοποίηση της προτεινόμενης μεθόδου που πραγματοποιήσαμε, και τέλος στο Κεφάλαιο 5 θα συνοψίσουμε παραθέτοντας τα συμπεράσματα της εργασίας και αναφέροντας μελλοντικές δυνατότητες έρευνας και ανάπτυξης στο συγκεκριμένο τομέα.

Κεφάλαιο 2

Παρούσα κατάσταση

Στο κεφάλαιο που ακολουθεί θα συγκεντρώσουμε και θα αναλύσουμε τα κριτήρια τα οποία είναι αναγκαία για την πέρα από κάθε αμφισβήτηση συλλογή και πιστοποίηση του περιεχομένου ιστοσελίδων, όπως αυτά προκύπτουν από τη βιβλιογραφία. Εν συνεχεία, για κάθε μία από τις 4 υπάρχουσες μεθόδους που έχουν επικρατήσει, θα κάνουμε αναφορά στη λειτουργικότητά τους και θα δούμε κατά πόσο πληρούν τα τεθέντα κριτήρια.

2.1 Κριτήρια

Από τη βιβλιογραφία που παραθέσαμε, καθώς και από τις προκλήσεις που παρουσιάζονται κατά τη χρήση των υφιστάμενων μεθόδων, έχουμε αναγνωρίσει εφτά κριτήρια, ο συνδυασμός των οποίων συμβάλλει στην πέρα από κάθε αμφισβήτηση συλλογή και πιστοποίηση του περιεχομένου ιστοσελίδων.

Όπως αναφέρθηκε, όσες περισσότερες και ακριβέστερες μέθοδοι χρησιμοποιούνται για τη συλλογή, αποθήκευση, χρονοσήμανση και πιστοποίηση του περιεχομένου, τόσες περισσότερες πιθανότητες υπάρχουν το δικαστήριο να αποδεχτεί την γνησιότητα του πειστηρίου αυτού [11].

Αναλυτικά τα κριτήρια που προαναφέραμε είναι τα παρακάτω:

- **Μη-τοπικότητα:** συνίσταται στην μη αποθήκευση του περιεχόμενου τοπικά, από τον ενδιαφερόμενο, και κατά συνέπεια την αποφυγή αμφισβήτησης είτε το περιεχομένου του ίδιου (και του αν επήλθε σε αυτό κάποια εσκεμμένη αλλοίωση), ή της τεχνικής επάρκειας του εκτελούντα την αποθήκευση.

- **Χρονοσήμανση:** αναφέρεται στο αν εφαρμόζονται μέθοδοι προσάρτησης της ημερομηνίας και ώρας στο υπό εξέταση πειστήριο, έτσι ώστε να αποδεικνύεται πέρα από κάθε αμφιβολία ο χρόνος αποθήκευσής του, και κατά συνέπεια η ημεροχρονολογία ύπαρξης του περιεχομένου στην ιστοσελίδα.

- **Περιεκτικότητα:** αναφέρεται στο αν ταυτόχρονα με το περιεχόμενο της ιστοσελίδας αποθηκεύεται και ο κώδικας της (HTML, css, javascript κλπ.), γεγονός που δίνει περισσότερες δυνατότητες εμβάθυνσης στο τι εμφάνιζε η ιστοσελίδα και πως το εμφάνιζε.

- **Αυθεντικότητα:** συνίσταται στο αν η μέθοδος διασφαλίζει, με τη χρήση ψηφιακών υπογραφών ή με άλλο τρόπο, το ποιος αποθήκευσε το περιεχόμενο.

- **Σφαιρικότητα:** αναφέρεται στο αν το περιεχόμενο της ίδιας ιστοσελίδας αποθηκεύεται ταυτόχρονα από πολλές τοποθεσίες, έτσι ώστε να καταδεικνύεται αν αυτό παρουσιαζόταν το ίδιο ή με τον ίδιο τρόπο σε διαφορετικά σημεία του κόσμου.

- **Αμεσότητα:** αναφέρεται στο αν η υπό εξέταση μέθοδος είναι άμεσα διαθέσιμη, δηλαδή την ώρα που εντοπίζεται το επίμαχο περιεχόμενο, ή αν χρειάζεται κάποια άλλη χρονοβόρα διαδικασία προκειμένου να εφαρμοστεί.

- **Ιδιωτικότητα:** συνίσταται στο αν η υπό εξέταση μέθοδος έχει διαθέσιμες λειτουργίες διασφάλισης της ιδιωτικότητας του χρήστη, κατά τη συλλογή και αποθήκευση των πειστηρίων.

Στις επόμενες ενότητες θα περιγράψουμε τις 4 υπάρχουσες μεθόδους, και θα δούμε ποια από τα κριτήρια που θέσαμε ικανοποιούν. Για κάθε μία από αυτές, θα παραθέσουμε έναν πίνακα όμοιο με το Σχήμα 2.1, όπου σε κάθε κριτήριο θα υπάρχει το σύμβολο ✓ αν η μέθοδος αυτή το πληρεί απόλυτα, και αντίστοιχα το σύμβολο ✗ αν δεν το πληρεί.

2.2 Τοπική αποθήκευση

Η πιο απλή και συχνότερα χρησιμοποιούμενη μέθοδος, όπως αναγράφεται και στο [16], είναι αυτή της τοπικής αποθήκευσης. Με τη μέθοδο αυτή ο ενδιαφερόμενος, μόλις εντοπίσει κάποιο περιεχόμενο που επιθυμεί να χρησιμοποιήσει σε

Μέθοδος \ Κριτήριο	Μη-Τοπικότητα	Χρονοσήμανση	Περιεκτικότητα	Αυθεντικότητα	Σφαιρικότητα	Αμεσότητα	Ιδιωτικότητα
Εξεταζόμενη μέθοδος	✓/✗	✓/✗	✓/✗	✓/✗	✓/✗	✓/✗	✓/✗

Σχήμα 2.1: Πρότυπος πίνακας πλήρωσης κριτηρίων

κάποια νομική διαδικασία, τότε το αποθηκεύει επί τόπου, χρησιμοποιώντας την επιλογή "Save As" που παρέχουν όλοι οι broswers. Το αποθηκευμένο περιεχόμενο δείχνει ακριβώς όπως η κανονική ιστοσελίδα, μόνο που κάποια στοιχεία στον κώδικα έχουν αλλάξει έτσι ώστε να μπορεί να προβάλεται από το δίσκο που έχει αποθηκευτεί.

Η μέθοδος της τοπικής αποθήκευσης έχει αρχικά το πλεονέκτημα της αμεσότητας, καθότι είναι απευθείας διαθέσιμη στο χρήστη, και δεν χρειάζονται ειδικές γνώσεις ούτε κάποιο επιπλέον κόστος για την αποθήκευση του περιεχομένου. Επίσης ένα σημαντικό πλεονέκτημα είναι η περιεκτικότητα του αποθηκευμένου περιεχομένου, καθώς αυτό συνήθως είναι πλήρους κώδικα, και περιέχει τόσο το κείμενο όσο και τη μορφοποίηση αυτού.

Σε ότι αφορά τα υπόλοιπα κριτήρια, σίγουρα αυτό της μη-τοπικότητας απουσιάζει. Αυτό σημαίνει ότι η τοπική αποθήκευση που κάνει ο χρήστης είναι αμφισβητήσιμη [16] και μπορεί εύκολα κάποιος να ισχυριστεί ότι ένα τοπικό αντίγραφο μπορεί να τροποποιηθεί με μεγάλη ευκολία. Και όντως αυτό μπορεί να συμβεί πολύ απλά, ανοίγοντας το περιεχόμενο με έναν editor, τροποποιώντας το, και αποθηκεύοντάς το πάλι.

Επίσης, σε ότι αφορά τη χρονοσήμανση, είναι γεγονός ότι τα αποθηκευμένα αρχεία θα εμπεριέχουν ημερομηνία και ώρα τροποποίησης. Αυτή όμως αντιστοιχεί στην ώρα του υπολογιστή, η οποία μπορεί να τροποποιηθεί πολύ εύκολα. Επομένως η μέθοδος αυτή δεν παρέχει ασφαλή και αδιαμφησβήτητη χρονοσήμανση.

Η έλλειψη αυθεντικότητας είναι ακόμη ένα σημαντικό μειονέκτημα της συγκεκριμένης μεθόδου. Πιο συγκεκριμένα, δεν προκύπτει με τρόπο αδιαμφηβήτητο το ποιος αποθήκευσε το περιεχόμενο, και κατά συνέπεια δημιούργησε το πειστήριο. Και πάλι υπάρχουν ενδείξεις από τα μεταδεδομένα των αρχείων, αλλά

σίγουρα δεν προσδιορίζουν μοναδικά το δημιουργό.

Τέλος, η εξεταζόμενη μέθοδος «αποδεικνύει» μόνο ποιο ήταν το περιεχόμενο μίας ιστοσελίδας, ειδωμένο από τον υπολογιστή του χρήστη. Δεν μας δίδεται καμία πληροφορία σχετικά με το ποιο περιεχόμενο ήταν την ίδια στιγμή διαθέσιμο σε άλλες τοποθεσίες ανά τον κόσμο.

Η παραπάνω αξιολόγηση της μεθόδου τοπικής αποθήκευσης φαίνεται συνοπτικά στο Σχήμα 2.2.

Μέθοδος \ **Κριτήριο**	Μη-Τοπικότητα	Χρονοσημανση	Περιεκτικότητα	Αυθεντικότητα	Σφιχτοφάγκωτο	Αμεσότητα	Ιδιωτικότητα
Τοπική αποθήκευση	✗	✗	✓	✗	✗	✓	✗

Σχήμα 2.2: Πλήρωση κριτηρίων για τη μέθοδο τοπικής αποθήκευσης

2.3 Μαγνητοσκόπηση πλοήγησης

Μία επίσης συχνά χρησιμοποιούμενη μέθοδος είναι αυτή της μαγνητοσκόπησης πλοήγησης [32]. Με τη μέθοδο αυτή, ο ενδιαφερόμενος αρχίζει να μαγνητοσκοπεί τα περιεχόμενα της οθόνης, με τη χρήση κατάλληλου λογισμικού, και προβαίνει διαδοχικά σε διάφορες κινήσεις ώστε: α) να δείξει τί ώρα είναι, είτε μέσω του ρολογιού του συστήματος, είτε από έγκριτες διαδικτυακές υπηρεσίες, και β) να δείξει ποιο είναι το περιεχόμενο συγκεκριμένης ιστοσελίδας, τη δεδομένη χρονική στιγμή.

Πολλές φορές, ταυτόχρονα με τον browser, ο χρήστης έχει ανοικτό κάποιο παράθυρο ώστε να προβάλει την εικόνα του εαυτού του από την κάμερα (ώστε να πιστοποιείται η ταυτοπροσωπία) αλλά και κάποιο παράθυρο με τη δικτυακή κίνηση (network traffic), ώστε να εμφαίνονται τα πακέτα που ανταλάσσονται μεταξύ πηγής και προορισμού.

Η μέθοδος της μαγνητοσκόπησης είναι γενικά αποδεκτή, διότι πληρεί τα περισσότερα από τα κριτήρια που έχουμε θέσει. Πρώτα απ΄ όλα ταυτοποιεί με σχεδόν αδιάψευστο τρόπο τόσο αυτόν που έκανε συλλογή του περιεχόμενου,

όσο και την ημεροχρονολογία που έλαβε χώρα η πράξη. Επιπλέον η μέθοδος αυτή είναι άμεσα διαθέσιμη, και μπορεί να επιπλέον να καταγράψει και προσωποποιημένο περιεχόμενο, αν ο χρήστης χρησιμοποιήσει τα στοιχεία πρόσβασης για να εισέλθει σε κάποια υπηρεσία.

Το μειονέκτημα της προαναφερόμενης μεθόδου είναι η μη ικανοποίηση του κριτηρίου της μη-τοπικότητας. Πιο συγκεκριμένα, η προβολή και αποθήκευση του περιεχομένου γίνεται από τον ίδιο τον ενδιαφερόμενο, με αποτέλεσμα να είναι εύκολο να αμφισβητηθεί η αντικειμενικότητά του.

Επιπρόσθετα, η μέθοδος υστερεί σημαντικά στο κομμάτι της διαφύλαξης της ιδιωτικότητας. Αυτό συμβαίνει διότι ο χρήστης δεν μπορεί να εξαιρέσει από την μαγνητοσκόπηση τα κομμάτια της ιστοσελίδας που θέλει να παραμείνουν ιδιωτικά, ή αν το έκανε η επέμβαση αυτή θα συνιστούσε τροποποίηση του πειστηρίου.

Τέλος, η μέθοδος της μαγνητοσκόπησης πλοήγησης δεν διασφαλίζει την σφαιρικότητα, δηλαδή το πως φαινόταν η ιστοσελίδα από διάφορες τοποθεσίες του κόσμου. Και πάλι ο χρήστης μπορεί να χρησιμοποιήσει διάφορες τοποθεσίες του ιστού που παρέχουν ελεύθερους proxy servers, και να μαγνητοσκοπήσει την πράξη του αυτή, αλλά αφ΄ ενός δεν θα μπορούσε να σώσει το αποτέλεσμα, και αφ΄ ετέρου αυτό θα ήθελε αυξημένες τεχνικές γνώσεις από τη μεριά του ενδιαφερόμενου.

Η παραπάνω αξιολόγηση της μεθόδου μαγνητοσκόπησης πλοήγησης φαίνεται συνοπτικά στο Σχήμα 2.3.

Μέθοδος \ **Κριτήριο**	Μη-Τοπικότητα	Χρονολόγηση	Περιεκτικότητα	Αυθεντικότητα	Σφαιρικότητα	Αμεσότητα	Ιδιωτικότητα
Μαγνητοσκόπηση πλοήγησης	✗	✓	✓	✓	✗	✓	✗

Σχήμα 2.3: Πλήρωση κριτηρίων για τη μέθοδο της μαγνητοσκόπησης πλοήγησης

2.4 Μέθοδος χρήσης τρίτων υπηρεσιών

Η μέθοδος που χρησιμοποιείται συχνότερα από χρήστες που δεν έχουν την τεχνογνωσία να αποθηκεύσουν το περιεχόμενο μόνοι τους είναι η χρήση τρίτων υπηρεσιών για τη συλλογή και πιστοποίηση αυτού. Πιο συγκεκριμένα, ο ενδιαφερόμενος προστρέχει σε υπηρεσίες που είναι διαθέσιμες στο διαδίκτυο [2, 7, 34] και αναλαμβάνουν, έναντι αντιτίμου, να αποθηκεύσουν το περιεχόμενο και να πιστοποιήσουν τη γνησιότητητά του.

Στην περίπτωση αυτή, προφανώς καλύπτεται το κριτήριο της μη-τοπικότητας, καθώς το περιεχόμενο αποθηκεύεται από έναν ανεξάρτητο επαγγελματία ή φορέα, χωρίς την εμπλοκή του ενδιαφερομένου. Ομοίως σε ότι αφορά το περιεχόμενο, καλύπτεται επίσης πλήρως η απαίτηση αποθήκευσης του συνόλου αυτού, μαζί με τη μορφοποίησή του.

Επίσης ικανοποιείται και το κριτήριο της χρονοσήμανσης, καθώς το σύνολο σχεδόν των υπηρεσιών αυτών παρέχουν ικανοποιητική σύνδεση του αποθηκευμένου περιεχομένου με τη χρονική στιγμή της αποθήκευσής του, ενώ σχετικά με την αυθεντικότητα, οι τρίτες αυτές υπηρεσίες συνήθως χρησιμοποιούν ψηφιακή υπογραφή για την πιστοποίηση του περιεχόμενου.

Το πιο σημαντικό ίσως μειονέκτημα της μεθόδου χρήσης τρίτων υπηρεσιών για την πιστοποίηση ιστοσελίδων είναι το γεγονός ότι δεν είναι άμεσα διαθέσιμες. Αυτό οφείλεται στο ότι αυτές προσφέρονται έναντι αντιτίμου, και θα πρέπει ο ενδιαφερόμενος να κάνει εγγραφή στην υπηρεσία, να καταβάλει το αντίστοιχο τίμημα, να επικοινωνήσει το τι θέλει να αποθηκεύσει κλπ.

Επίσης, σημαντικό μειονέκτημα είναι και το γεγονός ότι αυτές οι υπηρεσίες συνήθως δεν αποθηκεύουν προσωποποιημένο περιεχόμενο. Όταν ο ενδιαφερόμενος επιθυμεί να αποθηκευτεί κάποιο περιεχόμενο που βλέπει κατόπιν αυθεντικοποίησής του σε κάποια υπηρεσία, τότε θα πρέπει να κοινοποιήσει τα στοιχεία πρόσβασης στην τρίτη υπηρεσία, γεγονός που παραβαίνει την ιδιωτικότητά του. Για τον ίδιο λόγο, σε καμία απο τις υπηρεσίες που ελέγξαμε δεν υπάρχει διαθέσιμη η επιλογή της ιδιωτικότητας, δηλαδή να δίδεται η δυνατότητα στο τρίτο άτομο να αποκρύπτει κάποια στοιχεία που θέλουμε προκειμένου να διασφαλίσουμε την ιδιωτικότητά μας.

Τέλος, σε καμία από τις τρίτες αυτές υπηρεσίες δεν παρέχεται η δυνατότητα πραγματοποίησης πολλαπλών αιτήσεων από απομακρυσμένες τοποθεσίες, προκειμένου να δειχτεί το πως εμφανιζόταν το περιεχόμενο μίας ιστοσελίδας σε μία δεδομένη χρονική στιγμή.

Η παραπάνω αξιολόγηση της μεθόδου χρήσης τρίτων υπηρεσιών φαίνεται συνοπτικά στο Σχήμα 2.4.

Μέθοδος \ Κριτήριο	Μη-Τοπικότητα	Χρονοσήμανση	Περιεκτικότητα	Αυθεντικότητα	Σφαιρικότητα	Αμεσότητα	Ιδιωτικότητα
Χρήση τρίτων υπηρεσιών	✓	✓	✓	✓	✗	✗	✗

Σχήμα 2.4: Πλήρωση κριτηρίων για τη μέθοδο χρήσης τρίτων υπηρεσιών

2.5 Μέθοδος χρήσης ιστοσελίδων διατήρησης περιεχομένου

Μία τελευταία μέθοδος που θα εξετάσουμε είναι αυτή της χρήσης ιστοσελίδων διατήρησης περιεχομένου (τύπου web archving). Οι υπηρεσίες αυτές ασχολούνται κυρίως με την ιστορική αποθήκευση περιεχομένου, είτε όλου του ιστού δωρεάν, όπως το Internet Archive [31], είτε συγκεκριμένων ιστοσελίδων επί πληρωμή [12, 24, 27].

Όπως είναι φανερό, σε ότι αφορά τις ιστοσελίδες που διατηρούν περιεχόμενο δωρεάν, αυτό θα πρέπει και πάλι να συλλεχθεί και να πιστοποιηθεί με τη χρήση μίας εκ των μεθόδων που είδαμε στις προηγούμενες παραγράφου. Αντίθετα στις περισσότερες επι πληρωμή υπηρεσίες διατήρησης περιεχομένου, μπορούμε να έχουμε όλες τις διαδικασίες πιστοποίησης και χρονοσήμανσης επιθυμούμε για το περιεχόμενο που θα επιλέξουμε να διαατηρήσουμε. Επομένως τα κριτήρια χρονοσήμανσης και πιστοποίησης καλύπτονται πλήρως.

Επίσης, ικανοποιείται το κριτήριο της μη τοπικότητας και της περιεκτικότητας, μιας και το περιεχόμενο διατηρείται από την ιστοσελίδα διατήρησης περιεχομένου απομακρυσμένα και σε μορφή HTML, προκειμένου να μπορεί να προβληθεί στον browser.

Παρ΄ όλα αυτά, όπως και στην περίπτωση χρήσης τρίτων υπηρεσιών, έτσι και στην μέθοδο χρήσης ιστοσελίδων διατήρησης περιεχομένου έχουμε το μειονέκτημα της μη αποθήκευσης προσωποποιημένου περιεχομένου και στην απουσία προστασίας της ιδιωτικότητας, για τους λόγους που αναφέραμε παραπάνω. Η αμεσότητα είναι επίσης ένα στοιχείο που απουσιάζει, μιας και οι διαδικασίες για τη χρήση των (εμπορικών) ιστοσελίδων διατήρησης περιεχομένου είναι χρονο-

βόρες και κοστοβόρες.

Η παραπάνω αξιολόγηση της μεθόδου χρήσης ιστοσελίδων διατήρησης περιεχομένου φαίνεται συνοπτικά στο Σχήμα 2.5.

Μέθοδος \ Κριτήριο	Μη-Τοπικότητα	Χρονοβόρος	Περιεκτικότητα	Ανθεντικότητα	Σφαιρικότητα	Αμεσότητα	Ιδιωτικότητα
Χρήση ιστοσελίδων διατήρησης περιεχομένων	✓	✓	✓	✓	✗	✗	✗

Σχήμα 2.5: Πλήρωση κριτηρίων για τη μέθοδο χρήσης ιστοσελίδων διατήρησης περιεχομένου

Κεφάλαιο 3

Προτεινόμενη λύση: ProCAVE

Στο κεφάλαιο αυτό θα παρουσιάσουμε την κεντρική ιδέα πίσω από τη λύση που προτείνουμε για την αξιόπιστη αποθήκευση και πιστοποίηση του περιεχομένου ιστοσελίδων. Στο πλαίσιο αυτό, θα προτείνουμε μία συγκεκριμένη αρχιτεκτονική, θα αναλύσουμε τις επιμέρους λειτουργίες της και θα δούμε κατά πόσο αυτή η λύση ικανοποιεί τα τεθέντα κριτήρια. Τέλος θα παραθέσουμε μία σύγκριση με τις υπάρχουσες μεθόδους και θα κάνουμε διάφορες επισημάνσεις, κυρίως σχετικά με τις δυνατότητες anti-forensics (εφ' εξής αντίμετρα) και την αντιμετώπισή τους.

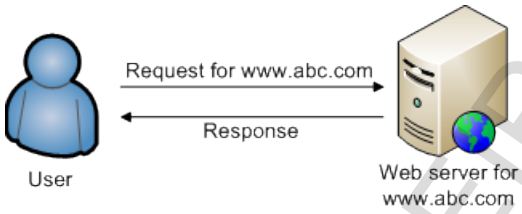
3.1 Κεντρική ιδέα

Για να αναλύσουμε την προτεινόμενη λύση, ας παραθέσουμε αρχικά το ίδιο σενάριο που χρησιμοποιήσαμε και στην ανάλυση των υπαρχουσών μεθόδων:

Σενάριο. *Σε μία ιστοσελίδα, έστω* `www.abc.com`*, βρίσκεται συγκεκριμένο περιεχόμενο το οποίο πρέπει να χρησιμοποιηθεί σε κάποια αποδεικτική διαδικασία. Για τον λόγο αυτό, πρέπει να αποθηκευτεί και να πιστοποιηθεί η γνησιότητά του, καθώς και ο χρόνος αποθήκευσης.*

Ιδέα. *Να χρησιμοποιηθεί μία ενδιάμεση ιστοσελίδα, έστω* `www.xyz.com`*. Έτσι όταν ο χρήστης εντοπίσει το επίμαχο περιεχόμενο στην ιστοσελίδα* `www.abc.com`*, να πλοηγηθεί στην ιστοσελίδα* `www.xyz.com/www.abc.com`*, η οποία θα αποθηκεύσει το επίμαχο περιεχόμενο τοπικά, και ταυτόχρονα θα το παρουσιάσει στον browser του χρήστη. Όταν ο χρήστης το επιλέξει, να μπορεί να αντλήσει μέσω της ενδιάμεσης ιστοσελίδας το επίμαχο περιεχόμενο σε μία μορφή που θα είναι πιστοποιημένη και χρονοσημασμένη.*

Η ιδέα που παραθέσαμε είναι πολύ απλή και στηρίζεται στις βασικές αρχές του ιστού. Όταν ένας χρήστης αιτείται, μέσω του browser, το περιεχόμενο μίας ιστοσελίδας, τότε το περιεχόμενο επιστρέφει από τον εξυπηρετητή (web server) με τη μορφή HTML, και παρουσιάζεται στον browser, όπως φαίνεται και στο Σχήμα 3.1.



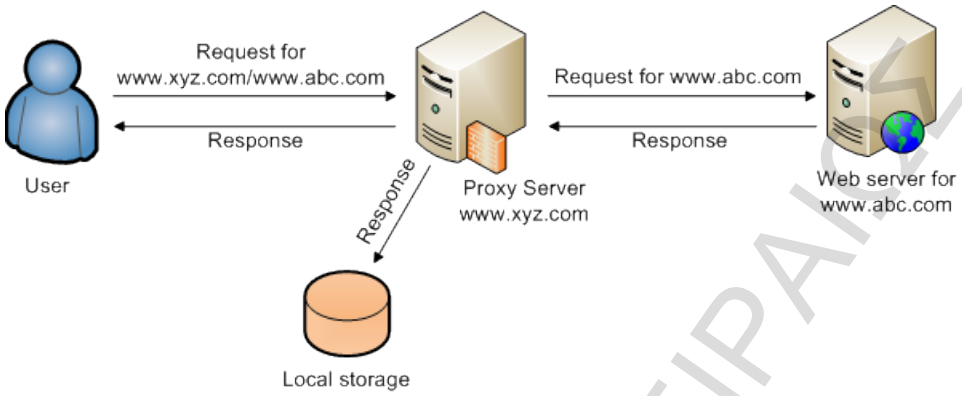
Σχήμα 3.1: Απλή αίτηση HTTP

Έτσι, στην ουσία αυτό που βλέπει ο χρήστης είναι το περιεχόμενο που του έχει επιστραφεί από τον εξυπηρετητή την δεδομένη χρονική στιγμή, και το οποίο έχει μορφοποιηθεί και παρουσιαστεί καταλλήλως από τον browser. Αυτό σημαίνει ότι όποια αποθήκευση, χρονοσήμανση και υπογραφή πραγματοποιηθεί από την μεριά του χρήστη, θα είναι ουσιαστικά στο περιεχόμενο το οποίο έχει επιστραφεί στον υπολογιστή του. Η προαναφερθείσα ιδιότητα αποτελεί και ένα απο τα σημαντικότερα μειονεκτήματα, όπως έχουμε δει, των υπαρχουσών μεθόδων.

Αντίθετα, στην λύση που προτείνουμε, ο χρήστης πραγματοποιεί το αίτημα μέσω της ιστοσελίδας `www.xyz.com`, η οποία εν συνεχεία μεταβιβάζει το αίτημα στον εξυπηρετητή της ιστοσελίδας `www.abc.com`. Το περιεχόμενο που επιστρέφεται, αποθηκεύεται τοπικά στον μηχανισμό της `www.xyz.com` και ταυτόχρονα παρουσιάζεται στον browser του χρήστη, όπως φαίνεται και στο Σχήμα 3.2.

Η λύση αυτή όπως είναι φανερό εξακολουθεί να δίνει την ίδια αίσθηση πλοήγησης στον χρήστη, αφού στον browser του παρουσιάζεται το ίδιο περιεχόμενο. Όμως, υπάρχουν οι εξής επιπλέον ιδιότητες:

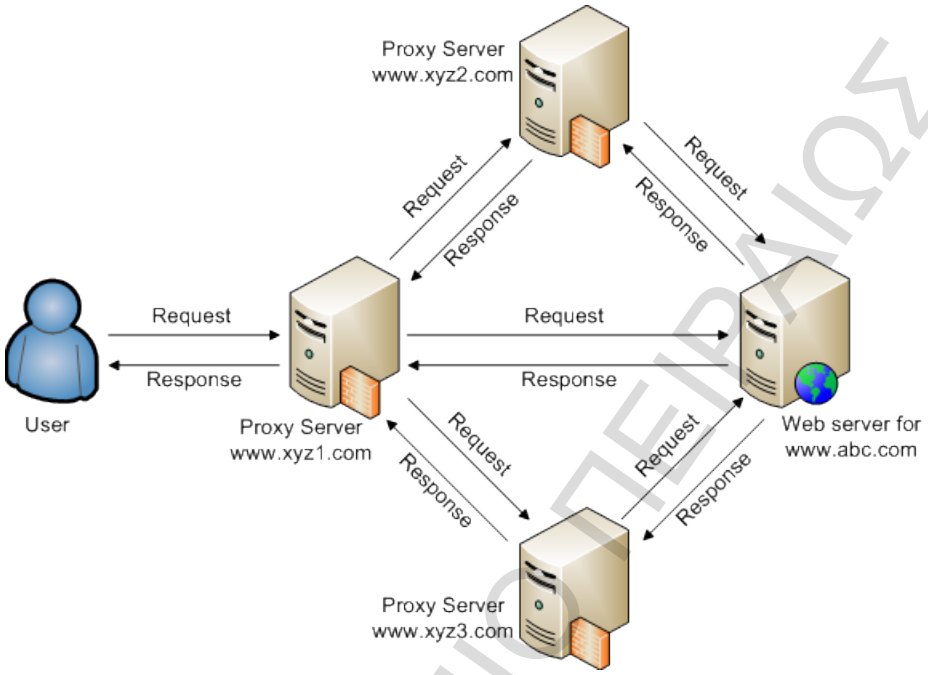
- Υπάρχει ένα κεντρικό σημείο (ανεξάρτητο από τον χρήστη) που είναι αποθηκευμένος ο κώδικας της ιστοσελίδας, απ΄ όπου μπορεί να ανακτηθεί σε δεύτερο χρόνο.
- Στο περιεχόμενο αυτό (στον αποθηκευμένο κώδικα δηλαδή) μπορούν να εφαρμοστούν αποδεκτές μέθοδοι πιστοποίησης, όπως ψηφιακή υπογραφή, καθώς και χρονοσήμανσης· και όλα αυτά από έμπιστες οντότητες.



Σχήμα 3.2: Αίτηση HTTP μέσω ενδιάμεσου εξυπηρετητή

Επιπλέον από τα προαναφερθέντα, η λύση που προτείνουμε έχει την δυνατότητα της παράλληλης εκτέλεσης από πολλές τοποθεσίες. Αυτό πρακτικά σημαίνει ότι η αίτηση `www.xyz.com/www.abc.com` δύναται να εκτελεστεί από πολλές διαφορετικές ενδιάμεσες ιστοσελίδες ταυτόχρονα, η κάθε μία από τις οποίες θα εκτελέσει όλες τις επιμέρους λειτουργίες ατομικά. Αυτό σημαίνει ότι η πολλαπλές αιτήσεις HTTP και η κατ΄ αντιστοιχία πολλαπλή αποθήκευση περιεχομένου θα δώσει την αίσθηση για το πως παρουσιαζόταν η ίδια ιστοσελίδα σε διαφορετικά σημεία την ίδια χρονική στιγμή. Μία γραφική απεικόνιση της λειτουργίας πολλαπλής εκτέλεσης δίδεται στο Σχήμα 3.3.

Τέλος, όπως προαναφέρθηκε, η προτεινόμενη λύση δίνει την ίδια αίσθηση πλοήγησης στον χρήστη, σαν να πραγματοποιούσε τις αιτήσεις HTTP μέσω της κανονικής οδού. Αυτό πρακτικά σημαίνει ότι αφ΄ ενός υπάρχει η δυνατότητα αποθήκευσης προσωποποιημένου (personalized) περιεχομένου του χρήστη, όπως π.χ. το προφίλ του σε μία ιστοσελίδα κοινωνικής δικτύωσης όπου απαιτείται κωδικός πρόσβασης. Αφ΄ ετέρου, η ύπαρξη προσωποποιημένου περιεχομένου που έχει αποθηκευτεί σε κάποιο κεντρικό σημείο δίνει τη δυνατότητα τροποποίησης του περιεχομένου αυτού με τέτοιο τρόπο ώστε να διασφαλίζεται η ιδιωτικότητα του χρήστη και να μην παρουσιάζονται δεδομένα που δεν θα ήθελε να παρουσιαστούν.



Σχήμα 3.3: Πολλαπλές αιτήσεις HTTP από διαφορετικές τοποθεσίες

3.2 Επιμέρους λειτουργίες

Η αρχιτεκτονική που έχουμε σχεδιάσει για την υλοποίηση της παραπάνω ιδέας ονομάζεται συνοπτικά ProCAVE (Privacy-preserving Collection and Authenticity Validation of online Evidence). Αποτελείται από ουσιαστικά δύο κομμάτια:

- Τον ενδιάμεσο μηχανισμό, που είναι της μορφής web proxy, και εφ΄ εξής θα αποκαλούμε Μηχανισμό Proxy.
- Τον μηχανισμό πιστοποίησης και χρονοσήμανσης (Collection and Validation of Authenticity), εφ΄ εξής Μηχανσμός CVA.

Στις υποενότητες που ακολουθούν θα παρουσιάσουμε συνοπτικά τους δύο αυτούς μηχανισμούς.

3.2.1 Μηχανισμός Proxy

Όπως προαναφέρθηκε, ο ενδιάμεσος αυτός μηχανισμός είναι τύπου web proxy[1]. Αυτό σημαίνει ότι λαμβάνει ένα αίτημα HTTP και πραγματοποιεί ένα αντίστοιχο άιτημα HTTP σε έναν απομακρυσμένο εξυπηρετητή. Όταν λάβει απάντηση (response), τότε την επιστρέφει στον πελάτη που είχε κάνει την αρχική αίτηση.

Στην περίπτωσή μας, όπως φαίνεται και στο Σχήμα 3.2 που έχουμε ήδη παραθέσει, η απάντηση από τον απομακρυσμένο εξυπηρετητή επιστρέφεται στον πελάτη, και ταυτόχρονα αποθηκεύεται τοπικά. Αυτό σημαίνει ότι, οποτεδήποτε ο χρήστης το ζητήσει, και με δεδομένο ότι το αποθηκευμένο περιεχόμενο είναι αυτό που επιθυμεί, τότε μπορεί να προχωρήσει στο επόμενο βήμα, που είναι η *Συλλογή και Πιστοποίηση* του περιεχομένου (Collection and Validation of Authenticity), με τη χρήση του Μηχανισμού CVA.

3.2.2 Μηχανισμός CVA

Όπως προαναφέραμε, μόλις ο χρήστης έχει ολοκληρώσει ένα αίτημα μέσα από τον Μηχανισμό Proxy, του έχει επιστραφεί το περιεχόμενο της αιτούμενης ιστοσελίδας ως απάντηση (response) η οποία επιπροσθέτως έχει αποθηκευτεί τοπικά στον Μηχανισμό Proxy. Η απάντηση αυτή αναγνωρίζεται από έναν μοναδικό κωδικό (id) που της δίδεται από το σύστημα.

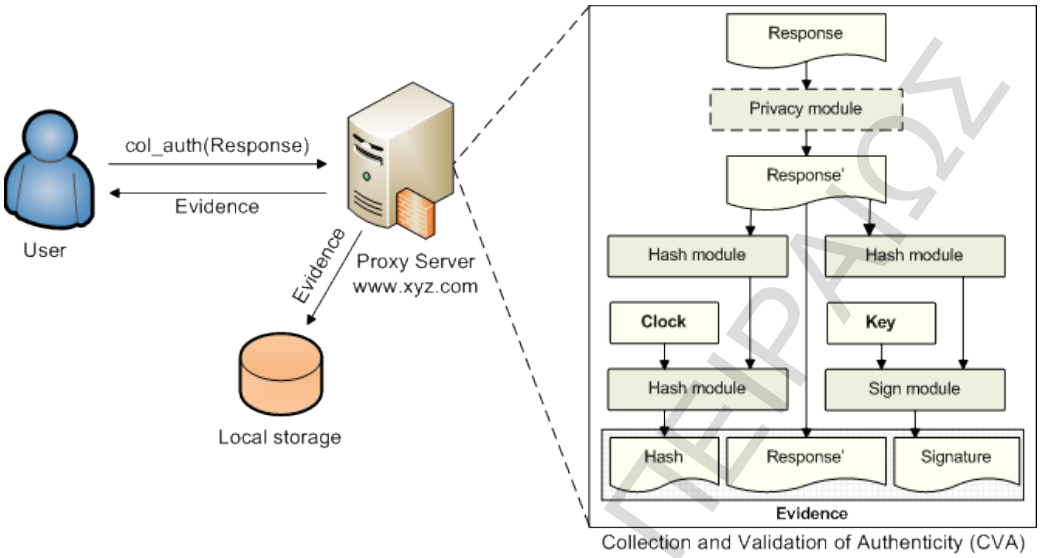
Τη στιγμή αυτή, μπορεί να προχωρήσει στην εφαρμογή κάποιων συναρτήσεων στο περιεχόμενο αυτό, προκειμένου να πιστοποιηθεί η κτήση του, να χρονοσημανθεί αλλά και να διατηρηθεί η ιδιωτικότητά του. Αυτό γίνεται, όπως φαίνεται στο Σχήμα 3.4, με τη βοήθεια του Μηχανισμού CVA.

Για να επιτευχθεί η χρήση του Μηχανισμού CVA, ακολουθούνται τα παρακάτω βήματα:

1. Ο χρήστης επιλέγει, εφόσον το επιθυμεί, μέσα από τον browser του τα κομμάτια της ιστοσελίδας που θέλει να αποκρύψει, προκειμένου να διατηρηθεί η ιδιωτικότητά του, και εν συνεχεία προχωρά στην επιλογή του Μηχανισμού CVA.

2. Ο μοναδικός κωδικός (id) της απάντησης αποστέλεται στον Μηχανισμό Proxy, μαζί με κάποια επιπλέον στοιχεία, που αντανακλούν το επίπεδο ιδιωτικότητας που θέλει να διατηρήσει ο χρήστης. Για να γίνει αυτό, υιοθετείται μία λογική whitelist/blacklist. Αυτό σημαίνει ότι ο χρήστης

[1] Σε άλλες περιπτώσεις στη βιβλιογραφία ο web proxy απαντάται ως HTTP proxy.



Σχήμα 3.4: Μηχανισμός Συλλογής και Πιστοποίησης (Collection and Validation of Authenticity Engine)

έχει τη δυνατότητα είτε να κρύψει κάποια κομμάτια και να διατηρήσει όλα τα άλλα ανέπαφα, είτε να διατηρήσει κάποια κομμάτια και να κρύψει όλα τα υπόλοιπα. Η απόκρυψη αυτή επιτυγχάνεται στέλνοντας στον Μηχανισμό Proxy τα HTML element ids, μαζί με την αντίστοιχη επιλογή (whitelist/blacklist).

3. Έχοντας τα προαναφερθέντα στοιχεία, ο Μηχανισμός CVA προχωράει στην ανάλογη τροποποίηση του περιεχόμενου, κρατώντας ένα αντίγραφο για λόγους αναφοράς. Η τροποποίηση αυτή εμπλέκει και το δημόσιο κλειδί του χρήστη, και αναπαριστάται στο Σχήμα 3.4 με διακεκομένη γραμμή λόγω της προαιρετικής χρήσης της.

4. Εν συνεχεία, το τροποποιημένο περιεχόμενο (Response') περνά μέσα από μία συνάρτηση hash, και παράγεται μία σύνοψη (digest). Στη σύνοψη αυτή προσαρτάται, σύμφωνα με το [13], η ημερομηνία και η ώρα αποθήκευσης του περιεχομένου, και η προκύπτουσα συμβολοσειρά περνά εκ νέου από μία συνάρτηση σύνοψης (digest), παράγοντας το πρώτο κομμάτι του πειστηρίου μας (evidence), που στο Σχήμα 3.4 αναφέρεται ως hash. *Το κομμάτι αυτό χρησιμοποιείται για την απόδειξη της ακριβής ημερομηνίας*

και ώρας ύπαρξης του περιεχομένου στην ιστοσελίδα.

5. Το δεύτερο κομμάτι του πειστηρίου είναι το τροποποιημένο περιεχόμενο αυτό καθε αυτό (Response'). *Το κομμάτι αυτό μπορεί να παρουσιαστεί ως έχει και χρησιμοποιείται για να παρουσιάσει ποιο ήταν το περιεχόμενο της ιστοσελίδας τη δεδομένη ημερομηνία και ώρα.*

6. Το τρίο και τελευταίο κομμάτι αποτελεί μία ψηφιακή υπογραφή της σύνοψης του τροποποιημένου περιεχομένου (Response'). Πιο συγκεκριμένα, το περιεχόμενο για άλλη μια φορά περνά μέσα μέσα από μία συνάρτηση hash, και παράγεται μία σύνοψη (digest). Η σύνοψη αυτή υπογράφεται με το ιδιωτικό κλειδί του Μηχανισμού CVA, και παράγεται αυτό που στο Σχήμα 3.4 αναπαρίσταται ως Signature. *Αυτό είναι ουσιαστικά η υπογραφή του Μηχανισμού CVA, και χρησιμοποιείται για να αποδειχθεί ποιος αποθήκευσε το συγκεκριμένο περιεχόμενο.*

7. Τελικά, όπως προείπαμε, τα τρία αυτά κομμάτια συνενώνονται για να παράξουν το πειστήριο (evidence) του περιεχομένου της ιστοσελίδας τη δεδομένη ημερομηνία και ώρα.

8. Τα παραπάνω βήματα ακολουθούνται και για το αρχικό (πριν τις αλλαγές) περιεχόμενο, το τελικό πειστήριο του οποίου αποθηκεύεται επίσης.

Αφού πραγματοποιηθούν τα βήματα της χρήσης του Μηχανισμού CVA, το πειστήριο του πραγματικού, καθώς και του τροποποιημένου περιεχομένου, επιστρέφεται στον χρήστη προκειμένου να χρησιμοποιηθεί σε οποιοαδήποτε διαδικασία, ενώ αντίγραφα αυτών θα διατηρηθούν στον Μηχανισμό Proxy για μελλοντική αναφορά, καθώς εύκολα μπορούν να αναζητηθούν με χρήση του μοναδικού κωδικού (id).

3.2.3 Πολλαπλές αιτήσεις HTTP

Όπως ήδη συζητήσαμε στην παρουσίαση της λύσης ProCAVE, ένα μοναδικό στιγμιότυπο του περιεχομένου μίας ιστοσελίδας, δηλαδή ένα πειστήριο, μπορεί να μην είναι αρκετό για την απόδειξη η μη μίας πράξης. Πολλές φορές στην πράξη χρειάζεται να αποδειχτεί το πως φαινόταν (παρουσιαζόταν) μία ιστοσελίδα σε διάφορα σημεία του κόσμου. Επιπλέον, μπορεί να καταστεί απαραίτητη η αποθήκευση του πειστηρίου αυτού σε διάφορες, γεωγραφικά διεσπαρμένες, τοποθεσίες.

Η σχεδίαση της προτεινόμενης λύσης ProCAVE ικανοποιεί αυτή την απαίτηση, όπως φαίνεται και στο Σχήμα 3.3. Στην περίπτωση αυτή, όταν ο Μηχανισμός Proxy λαμβάνει μία αίτηση HTTP, την διεκπεραιώνει αλλά ταυτόχρονα προωθεί ένα αντίγραφο αυτής σε άλλους web proxies. Ως συνέπεια, κάθε ένας από αυτούς θα συλλέξει και θα πιστοποιήσει το περιεχόμενο με τη δικό του Μηχανισμό CVA, όπως το περιεχόμενο αυτό παρουσιάζεται στην γεωγραφική τοποθεσία στην οποία βρίσκεται ο εξυπηρετητής. Μόλις πραγματοποιηθεί αυτό, ο κάθε ένας από τους web proxies θα επιστρέψει το πειστήριο του στον αιτούντα Μηχανισμό Proxy, ο οποίος θα αναλάβει να τα συγκεντρώσει και να τα επιστρέψει στον χρήστη.

Είναι φανερό ότι υπάρχει μεγάλη πιθανότητα τα πολλά διαφορετικά συλλεχθέντα πειστήρια να διαφέρουν μεταξύ τους. Παρ΄ όλα αυτά, η συλλογή αυτών των πειστηρίων, ειδικά από διαφορετικές τοποθεσίες, ικανοποιεί τις απαιτήσεις της: α) πολλαπλής αποθήκευσης για λόγους ασφαλείας, και β) παρουσίασης του πως φαινόταν το περιεχόμενο της υπό εξέταση ιστοσελίδας σε διάφορες τοποθεσίες ανά τον κόσμο. Στην επόμενη ενότητα θα αναλύσουμε το κατά πόσο η προτεινόμενη λύση ικανοποιεί τα κριτήρια που θέσαμε στην Ενότητα 2.1.

3.3 Πλήρωση κριτηρίων

Η λύση ProCAVE που προτείνουμε είναι πλήρως εναρμονισμένη με το σύνολο των τεθέντων κριτηρίων, ο συνδυασμός των οποίων εγγυάται αδιαμφησβήτητη συλλογή και πιστοποίηση περιεχομένου ιστοσελίδων. Πιο συγκεκριμένα:

Μη-τοπικότητα Από την περιγραφή της λύσης ProCAVE γίνεται φανερό ότι αποφεύγεται η παγίδα της τοπικότητας. Αυτό συμβαίνει διότι, αφ΄ ενός η αίτηση HTTP πραγματοποιείται μέσω του ενδιάμεσου web proxy που περιγράψαμε, και αφ΄ ετέρου ο κώδικας που εμπεριέχεται στην απάντηση (response) αποθηκεύεται απομακρυσμένα στον ενδιάμεσο εξυπηρετητή. Έτσι ο χρήστης σε κανένα σημείο δεν παρεμβαίνει για να αποθηκεύσει το περιεχόμενο της ιστοσελίδας· αντίθετα αυτό γίνεται κεντρικά από τον εξυπηρετητή μας, με αποτέλεσμα να μην μπορεί να αμφισβητηθεί.

Χρονοσήμανση Όπως είδαμε, η απάντηση (response) περνάει μέσα από τον Μηχανισμό CVA, όπου μεταξύ άλλων στη σύνοψη αυτής προσαρτάται, σύμφωνα με το [13], η ημερομηνία και η ώρα αποθήκευσης του περιεχομένου. Αυτό είναι και μία επαρκής απόδειξη για την ώρα που αποθηκεύτηκε το περιεχόμενο, και μπορεί να επαληθευτεί ανά πάσα στιγμή.

Περιεκτικότητα Το περιεχόμενο που αποθηκεύεται είναι της μορφής HTTP, και μπορεί επίσης να περιέχει και κάποια άλλα στοιχεία της ιστοσελίδας όπως φωτογραφίες, κώδικα άλλης μορφής (css,javascript) κλπ. Αυτό σημαίνει ότι όποιος λάβει αντίγραφο του πειστηρίου, μπορεί να ανοίξει το κεντρικό αρχείο HTML με έναν browser και να δει πως παρουσιαζόταν η ιστοσελίδα τη συγκεκριμένη ημερομηνία και ώρα αποθήκευσής της. Είναι φανερό ότι με αυτόν τον τρόπο είναι διαθέσιμος ο πλήρης κώδικας της ιστοσελίδας, και κάθε ενδιαφερόμενος μπορεί να δει μέσα σε αυτόν πολλές χρήσιμες πληροφορίες, π.χ. τη λειτουργία των scripts, κρυφούς συνδέσμους (hidden links) ή κρυμμένο κώδικα που υπάρχει στην ιστοσελίδα κλπ.

Αυθεντικότητα Όπως περιγράψαμε κατά την περιγραφή της λύσης, ο Μηχανισμός CVA παρέχει και λειτουργία «ψηφιακής υπογραφής» του περιεχομένου. Αυτό πραγματοποιείται με τη χρήση του ιδιωτικού κλειδιού του Μηχανισμού CVA, με το οποίο γίνεται η κρυπτογράφηση της σύνοψης του περιεχομένου. Έτσι ο ενδιαφερόμενος, χρησιμοποιώντας το δημόσιο κλειδί του Μηχανισμού CVA μπορεί να διαπιστώσει άμεσα αν η αποθήκευση του περιεχόμενου, και κατά συνέπεια η ψηφιακή υπογραφή αυτού, πραγματοποιήθηκαν από τον συγκεκριμένο μηχανισμό.

Σφαιρικότητα Η σφαιρικότητα όπως έχουμε περιγράψει είναι η ιδιότητα της συλλογής του περιεχομένου μίας ιστοσελίδας από πολλές διαφορετικές, γεωγραφικά διεσπαρμένες, τοποθεσίες. Αυτό είναι κάτι που ικανοποιείται πλήρως με τη λύση ProCAVE, καθώς όπως είδαμε η ίδια αίτηση μπορεί να σταλεί από τον Μηχανισμός Proxy σε πολλούς άλλους παρόμοιους εξυπηρετητές, και να λάβει τα αποτελέσματα. Τα αποτελέσματα αυτά θα είναι ευδιάκριτα μεταξύ τους, καθώς θα έχουν υπογραφεί με διαφορετικο κλειδί, και θα αποδεικνύουν πως παρουσιαζόταν μία ιστοσελίδα σε συγκεκριμένη ημερομηνία και ώρα σε διάφορες τοποθεσίες. Ο ρόλος της λύσης ProCAVE είναι απλώς να συνενώσει αυτές τις απαντήσεις σε ένα ενιαίο πειστήριο, που θα επιστραφεί στον χρήστη όταν επιλέξει τη χρήση του Μηχανισμού CVA.

Αμεσότητα Σε ότι αφορά την αμεσότητα, δηλαδή το κατά πόσο η λύση μας θα είναι άμεσα διαθέσιμη στο χρήστη, αυτό είναι ένα από τα βασικά πλεονεκτήματα του ProCAVE. Όπως είδαμε, στον χρήστη δίνεται άμεσα η δυνατότητα να αποθηκεύσει με ασφαλή και εύκολο τρόπο τόσο δημόσιο, όσο και προσωποποιημένο (personalized) περιεχόμενο, χωρίς την εμπλοκή του σε περίπλοκες διαδικασίες. Επιπλέον, η αποθήκευση του περιεχόμενου στον Μηχανισμό Proxy είναι ανεξάρτητη από τη λειτουργία του Μηχανισμού Proxy. Έτσι, σε περίπτωση που η λύση ProCAVE αποτελέσει υπηρεσία επί πληρωμεί, η λειτουργία

της πλοήγησης και αποθήκευσης του περιεχομένου (δηλαδή η χρήση του Μηχανισμού Proxy) μπορεί να παρέχεται δωρεάν, ενώ η πιστοποίηση αυτού και η τελική απόδοσή του στον χρήστη (δηλαδή η χρήση του Μηχανισμού CVA) μπορεί να παρέχεται μόνο αν καταβληθεί το αντίστοιχο τίμημα.

Ιδιωτικότητα Τέλος, η λύση ProCAVE που προτείνουμε είναι η μόνη από τις υπάρχουσες λύσεις που δίνει τη δυνατότητα διατήρησης της ιδιωτικότητας του χρήστη. Με απλό τρόπο, χωρίς πολύπλοκες διαδικασίες, και με διατήρηση του αυθεντικού περιεχομένου για λόγους αναφοράς, ο χρήστης μπορεί να επιλέξει το περιεχόμενο που θέλει να αποκρύψει για λόγους ιδιωτικότητας και να προχωρήσει στη νόμιμη «αλλοίωση» αυτού.

Συνοπτικά, η ικανοποίηση των τεθέντων κριτηρίων από τη μέθοδο ProCAVE φαίνεται στον παρακάτω πίνακα:

Κριτήριο / Μέθοδος	Μη-Τοπικότητα	Χρονοσήμανση	Περιεκτικότητα	Αυθεντικότητα	Σφαιρικότητα	Αμεσότητα	Ιδιωτικότητα
ProCAVE	✓	✓	✓	✓	✓	✓	✓

Σχήμα 3.5: Πλήρωση κριτηρίων για τη μέθοδο ProCAVE

3.4 Σύγκριση με άλλες μεθόδους

Όπως είδαμε στην προηγούμενη παράγραφο, η λύση ProCAVE που προτείνουμε ικανοποιεί όλα τα τεθέντα κριτήρια για την ασφαλή και πέρα από κάθε αμφισβήτηση αποθήκευση του περιεχομένου ιστοσελίδων για χρήση του σε αποδεικτικές διαδικασίες. Η πρότασή αυτή υπερέχει σε σύγκριση με όλες τις υπάρχουσες μεθόδους, σε ένα ή περισσότερα σημεία, που θα εξετάσουμε ευθύς αμέσως.

Σε σύγκριση με τη μέθοδο της τοπικής αποθήκευσης, είναι φανερό ότι η λύση ProCAVE υπερτερεί στο ότι παρέχει δυνατότητες ακριβούς χρονοσήμανσης, σφαιρικότητας και αυθεντικότητας. Αυτά δεν μπορούν να επιτευχθούν

όταν ένας απλός χρήστης αποθηκεύει τοπικά μία ιστοσελίδα· ειδικά η επίτευξη της αυθεντικότητας είναι πρακτικά αδύνατη, γιατί θα πρέπει κάθε χρήστης να είναι κάτοχος ζεύγους κλειδιών.

Σε σύγκριση με τη μέθοδο της μαγνητοσκόπησης πλοήγησης, είναι σαφές ότι η λύση ProCAVE υπερτερεί διότι εξασφαλίζει την σφαιρικότητα και την διασφάλιση της ιδιωτικότητας· ειδικά το δεύτερο είναι αδύνατον να επιτευχθεί όταν μαγνητοσκοπείται η πλοήγηση, γιατί θα έπρεπε ο χρήστης να κάνει επεξεργασία του βίντεο και να «κρύψει» τα σημεία που θέλει να δισφαλίσει.

Σε σύγκριση με τη μέθοδο της χρήσης τρίτων υπηρεσιών, η λύση ProCAVE υπερτερεί κυρίως στην αμεσότητα, καθώς μπορεί να χρησιμοποιηθεί άμεσα από τον χρήστη, μόλις αυτός διαπιστώσει κάποιο περιεχόμενο που χρίζει αποθήκευσης. Αντίθετα, όπως έχουμε δει, οι τρίτες υπηρεσίες πρέπει να αγοραστούν, ή να γίνει κάποια εγγραφή, διαδικασίες δηλαδή που είναι χρονοβόρες, δεν μπορούν να γίνουν από όλους τους χρήστες, και επιπλέον δεν διασφαλίζουν την ιδιωτικότητα του χρήστη.

Σε σύγκριση με τη μέθοδο των ιστοσελίδων διατήρησης περιεχομένου, αυτές επίσης πρέπει να αγοραστούν και δεν διασφαλίζουν την ιδιωτικότητα του χρήστη, όμοια δηλαδή με τη χρήση τρίτων υπηρεσιών.

Συνοπτικά η σύγκριση όλων των μεθόδων δίδεται στο Σχήμα 3.6.

Μέθοδος \ Κριτήριο	Μη-Τοπικότητα	Χρονοσήμανση	Περιεκτικότητα	Αυθεντικότητα	Σφαιρικότητα	Αμεσότητα	Ιδιωτικότητα
Τοπική αποθήκευση	✗	✗	✓	✗	✗	✓	✗
Μαγνητοσκόπηση πλοήγησης	✗	✓	✓	✓	✗	✓	✗
Χρήση τρίτων υπηρεσιών	✓	✓	✓	✓	✗	✗	✗
Χρήση ιστοσελίδων διατήρησης περιεχομένων	✓	✓	✓	✓	✗	✗	✗
ProCAVE	✓	✓	✓	✓	✓	✓	✓

Σχήμα 3.6: Σύγκριση διαφόρων μεθόδων συλλογής περιεχομένου

Από όλα τα παρατεθέντα σε αυτό το κεφάλαιο στοιχεία, καθώς και από τη

σύγκριση που κάναμε, αποδεικνύεται ότι η μέθοδος ProCAVE που προτείνουμε φαίνεται να είναι η πληρέστερη και ασφαλέστερη λύση για την πέρα από κάθε αμφισβήτηση συλλογή και πιστοποίηση περιεχομένου ιστοσελίδων, αφού ικανοποιούνται όλα τα κριτήρια που θέσαμε και όλες οι προδιαγραφές που αναφέρονται στα [9, 19]. Τέλος, τα πειστήρια που δίνει η λύση μας μπορούν να θεωρηθούν και δημόσια έγγραφα, όπως αναφέρεται στο [20], αν το σύστημα ProCAVE λειτουργεί σε μία ορισμένη από τον νόμο δημόσια υπηρεσία.

Στην επόμενη ενότητα θα παρουσιάσουμε ένα πρωτότυπο της λύσης ProCAVE που υλοποιήσαμε, τον τρόπο ανάπτυξης καθώς και κάποια πειράματα που διεξάγαμε προκειμένου να ελέγξουμε την ορθότητα της πρότασής μας.

Κεφάλαιο 4

Υλοποίηση και πειραματική αξιολόγηση

Στο τέταρτο κεφάλαιο αυτής της εργασίας θα αναφερθούμε στην πιλοτική υλοποίηση του συστήματος ProCAVE που κάναμε, και κυρίως στον τρόπο ανάπτυξης και στα εργαλεία που χρησιμοποιήθηκαν. Επίσης θα αναφερθούμε στα αποτελέσματα χρήσης του ProCAVE με διάφορες ιστοσελίδες, και στο κατά πόσο αυτά αποδεικνύουν την ορθότητα της συγκεκριμένης λύσης. Τέλος θα κάνουμε διάφορες επισημάνσεις πάνω σε κάποια τεχνικά θέματα σχετικά με το σύστημα ProCAVE, και κυρίως με τεχνικές απόκρυψης δεδομένων (anti-forensics).

4.1 Τρόπος υλοποίησης

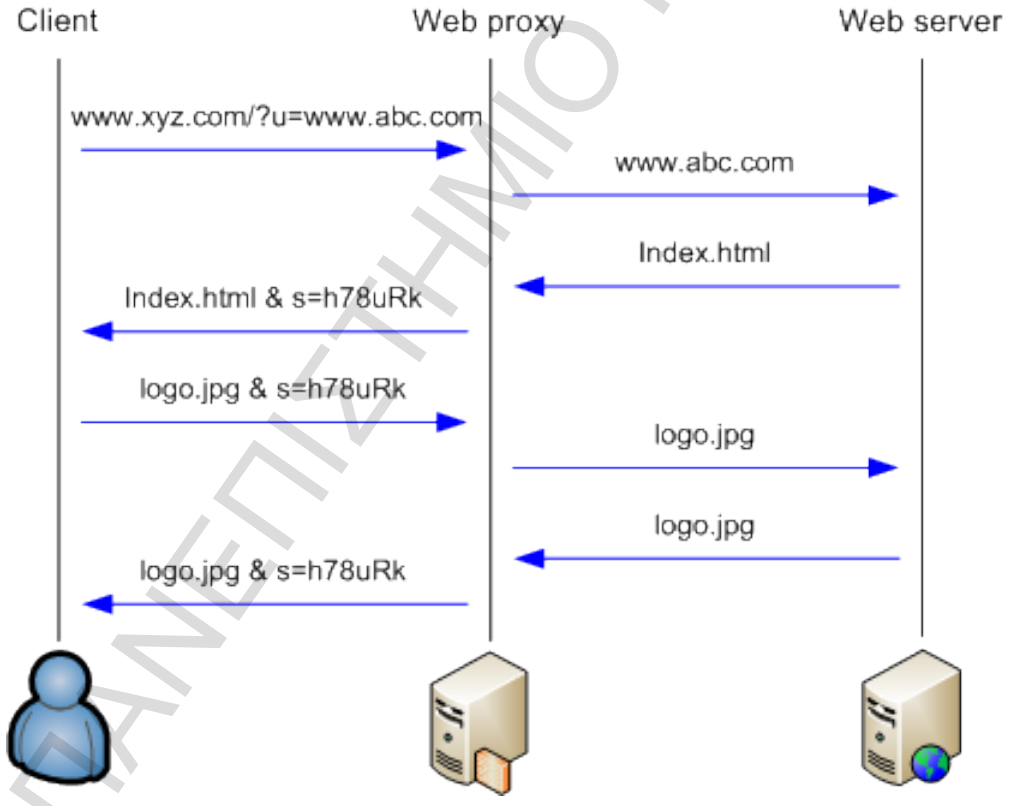
Η υλοποίηση της λύσης μας πραγματοποιήθηκε με τη γλώσσα προγραμματισμού PHP, και με τη χρήση του συστήματος διαχείρισης βάσεων δεδομένων MySQL [26]. Η ιδέα πίσω από την ανάπτυξη στηρίχθηκε στο παράδειγμα ενός απλού web proxy, που περιγράφεται στο [22], που χρησιμοποιήθηκε ως Μηχανισμός Proxy και ο οποίος επεκτάθηκε ώστε να συμπεριλάβει την υλοποίηση του Μηχανισμού CVA, που αναπτύξαμε στο προηγούμενο κεφάλαιο.

Πιο συγκεκριμένα, κάθε φορά που πραγματοποιείται μία αίτηση για μία ιστοσελίδα, έστω `www.abc.com`, μέσα από τον Μηχανισμό Proxy, τότε ο μηχανισμός αυτός αναλαμβάνει να μεταφέρει το αίτημα στον web server που εξυπηρετεί την ιστοσελίδα. Αυτός θα του επιστρέψει ένα κύριο αρχείο, έστω index.html. Το αρχείο αυτό, όπως έχουμε δει, θα αποθηκευτεί τοπικά στον Μηχανισμό

Proxy, και θα αποσταλεί επίσης στον χρήστη.

Επίσης, στην υλοποίηση που κάναμε, μαζί με το αρχείο index.html, ο Μηχανισμός Proxy θα επιστρέψει και ένα τυχαίο αλφαριθμητικό έξι ψηφίων. Αυτό συμβαίνει για μπορέσει να «συγκεντρώσει» όλες τις επόμενες αιτήσεις που θα γίνουν, και θα αφορούν την ίδια ιστοσελίδα, έτσι ώστε να παρουσιάσει ένα ενιαίο αποτέλεσμα.

Έτσι, αν επιστρέφοντας η ιστοσελίδα index.html απαιτηθεί να ληφθεί κάποια εικόνα, έστω logo.jpg που ήταν ορισμένη στον κώδικα, τότε θα πραγματοποιηθεί ένα request το οποίο όμως αυτή τη φορά θα έχει και το χαρακτηριστικό αλφαριθμητικό. Φυσικά και η εικόνα αυτή (καθώς και οτιδήποτε άλλο κατέβει και είναι σχετικό με την ιστοσελίδα) θα αποθηκευτεί τοπικά στον Μηχανισμό Proxy. Μία απεικόνιση αυτής της επικοινωνίας μεταξύ χρήστη, Μηχανισμού Proxy και τελικής ιστοσελίδας φαίνεται στο Σχήμα 4.1.



Σχήμα 4.1: Διαδικασία λειτουργίας της υλοποίησης ProCAVE

Τώρα όταν ο χρήστης επιλέξει την αποθήκευση της ιστοσελίδας (και κατά συνέπεια και των συνδεδεμένων με αυτήν αρχείων), τότε θα τρέξει ο Μηχανισμός CVA που έχουμε υλοποιήσει. Ο μηχανισμός αυτός εκτελεί όλες τις διαδικασίες (σύνοψης, χρονοσήμανσης, υπογραφής κλπ) εξ' ολοκλήρου για κάθε ένα από τα αρχεία χωριστά. Εν συνεχεία χρησιμοποιεί το εξαψήφιο αλφαριθμητικό για να συνενώσει όλα τα σχετικά αρχεία σε ένα πειστήριο.

Σε επίπεδο χρονοσήμανσης, χρησιμοποιήθηκε η έκδοση 320 bit του αλγορίθμου RIPEMD για σύνοψη (hashing), και συγκεκριμένα με τη βοήθεια της συνάρτησης hash_file() της PHP, ενώ για την ημερομηνία και ώρα χρησιμοποιήθηκαν οι βασικές συναρτήσεις της PHP. Σε ότι αφορά την ψηφιακή υπογραφή, χρησιμοποιήθηκε ο αλγόριθμος SHA-1 για σύνοψη (hashing), ακολουθούμενος από κρυπτογράφηση με τη βοήθεια του ιδιωτικού κλειδιού του web proxy το οποίο δημιουργήσαμε με χρήση της OpenSSL βιβλιοθήκης [23].

4.2 Λειτουργία του συστήματος

Αφού στην προηγούμενη ενότητα αναφερθήκαμε στη σχεδίαση και στην υλοποίηση του συστήματος ProCAVE, στην παρούσα θα κάνουμε μία σύντομη αναφορά στη λειτουργικότητα αυτού, και πιο συγκεκριμένα πως επιτυγχάνεται η αποθήκευση του περιεχομένου που μας ενδιαφέρει κάθε φορά.

Αρχικά επισκεπτόμαστε την ιστοσελίδα `www.procave.gr`, όπου έχουμε εγκατεστημένο το σύστημά μας. Εκεί μας εμφανίζεται η πρώτη σελίδα του συστήματος, με το λογότυπο αυτής, και με ένα πεδίο όπου μπορούμε να βάλουμε τη διεύθυνση που θέλουμε να πλοηγηθούμε, όπως φαίνεται και στο Σχήμα 4.2.

Σχήμα 4.2: Αρχική σελίδα του ProCAVE

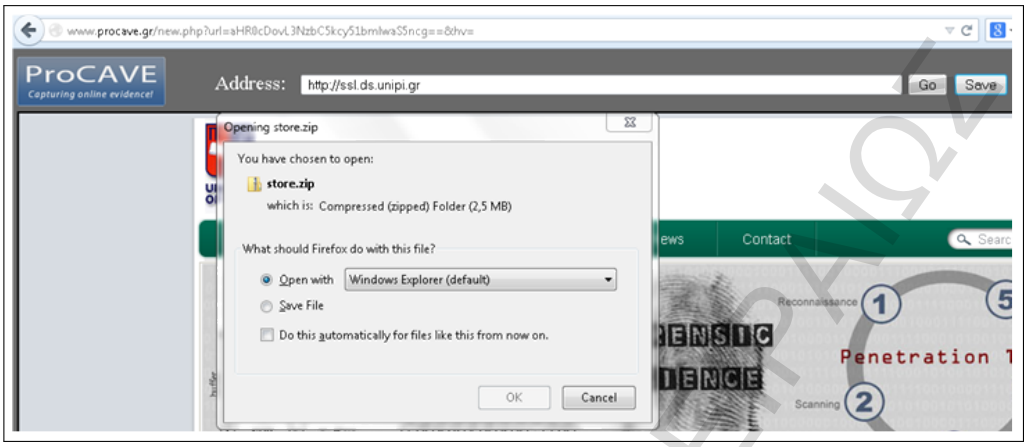
Αφού γράψουμε την επιθυμητή διεύθυνση URL στο πεδίο “Address”, πατάμε το κουμπί “Go”. Δίνοντας αυτήν την επιλογή, αμέσως πραγματοποιούνται τα βήματα που περιγράψαμε σε προηγούμενες ενότητες, και στην οθόνη μας εμφανίζεται το περιεχόμενο της ιστοσελίδας που έχουμε επιλέξει, όπως φαίνεται και στο Σχήμα 4.3.

Σχήμα 4.3: Πλοήγηση μέσα από το ProCAVE

Τώρα, σύμφωνα με τις προδιαγραφές που έχουμε αναπτύξει, το περιεχόμενο που έχει παρουσιαστεί στην οθόνη μας έχει αποθηκευτεί και τοπικά στον Μηχανισμό Proxy. Τώρα, και εφόσον αποφασίσουμε ότι το περιεχόμενο της ιστοσελίδας μας ενδιαφέρει να αποθηκευτεί για αποδεικτικούς λόγους, πατάμε το κουμπί “Save”. Αμέσως πραγματοποιείται η εκτέλεση του Μηχανισμού CVA, γίνονται άμεσα όλα τα απαραίτητα βήματα σύνοψης, χρονοσήμανσης και υπογραφής, και μας παρουσιάζεται ένα παράθυρο διαλόγου προκειμένου να σώσουμε το πειστήριο τοπικά. Η τελική αυτή οθόνη φαίνεται στο Σχήμα 4.4.

4.3 Αποτελέσματα πειραμάτων

Η παραπάνω λειτουργικότητα δοκιμάστηκε με πολλές ιστοσελίδες διαφορετικού περιεχομένου (ειδησεογραφικές, ακαδημαϊκές, ψυχαγωγίας). Για κάθε μία από αυτές τις ιστοσελίδες, το περιεχόμενο τους συλλέχθηκε με το εργαλείο που υλοποιήσαμε, και στη συνέχεια έγινε σύγκριση αυτού με το περιεχόμενο που προέρχεται από τη διαδικασία “Save As” που παρέχεται από όλους τους browsers. Η σύγκριση πραγματοποιήθηκε με τη χρήση του Similarity Analyzer [28],



Σχήμα 4.4: Αποθήκευση του περιεχομένου του ProCAVE

που εξετάζει το περιεχόμενο δύο ιστοσελίδων και υπολογίζει το ποσοστό % της ομοιότητάς τους σε επίπεδο κώδικα HTML και κειμένου. Τα αποτελέσματα για 10 γνωστές ιστοσελίδες φαίνονται στο Σχήμα 4.5.

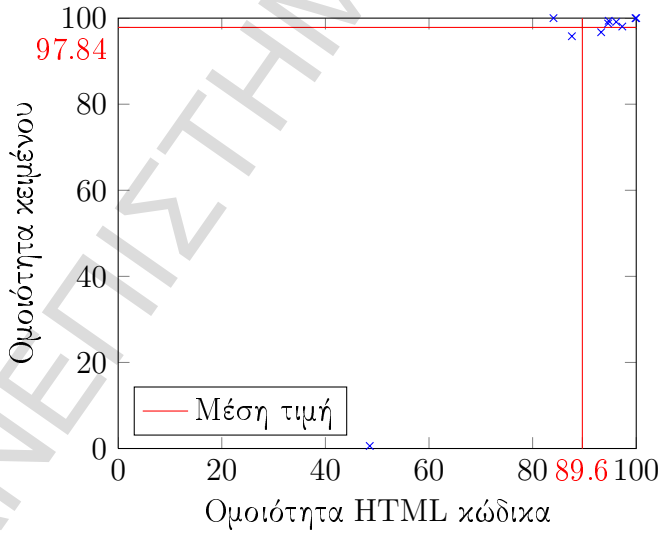
Από τη μελέτη του σχήματος συμπεραίνουμε ότι, στις περισσότερες των περιπτώσεων, η ομοιότητα HTML είναι κοντά στο 90%. Αυτό είναι αναμενόμενο αποτέλεσμα, διότι τόσο η μέθοδος ProCAVE όσο και η αποθήκευση με χρήση της διαδικασίας "Save As" προκαλούν αλλαγές στον κώδικα που αποθηκεύεται, έτσι ώστε οι σύνδεσμοι (links) που εμπεριέχονται σε αυτόν να δείχνουν πλέον σε τοπικούς δίσκους, και οι αλλαγές αυτές γίνονται με διαφορετικό τρόπο σε κάθε μέθοδο.

Από την άλλη, η ομοιότητα κειμένου είναι σχεδόν στο 98%, που σημαίνει ότι αυτό που αποθηκεύεται με τη χρήση του ProCAVE είναι πολύ κοντά με αυτό που βλέπει ο χρήστης στον browser του. Η διαφορά του 2% κατά μέσο όρο οφείλεται κυρίως σε κείμενο το οποίο αλλάζει ασύγχρονα, διαφημίσεις, ή άλλα στοιχεία τα οποία τυγχάνουν να είναι διαφορετικά από στιγμή σε στιγμή σε μία ιστοσελίδα.

Μία γραφική αναπαράσταση των αποτελεσμάτων φαίνεται στο Σχήμα 4.6 που ακολουθεί.

	Ομοιότητα	
Ιστοσελίδα	**HTML**	**Κείμενο**
inria.fr	99,86%	99,98%
ssl.ds.unipi.gr	100%	100%
news.yahoo.com	93,23%	96,72%
spiegel.de	94,68%	99,38%
behind-the-enemy-lines.com	48,56%	90,59%
maawg.org	94,57%	98,75%
ansa.it	97,31%	98,05%
bbc.co.uk	87,58%	95,79%
slashdot.org	96,14%	99,13%
xkcd.xom	84,03%	100%

Σχήμα 4.5: Αποτελέσματα σύγκρισης του περιεχομένου των ιστοσελίδων



Σχήμα 4.6: Γραφική αναπαράσταση αποτελεσμάτων σύγκρισης

4.4 Επισημάνσεις

Όπως επισημαίνεται και στο [1], από την πρώτη μέρα που εμφανίστηκαν οι ηλεκτρονικοί υπολογιστές και τα δίκτυα, ένα σημαντικό θέμα ήταν η απόκρυψη δεδομένων από τα εργαλεία και τις μεθόδους εγκληματολογικής ανάλυσης (forensics).

Η λύση ProCAVE που προτείνουμε, σαν εργαλείο συλλογής περιεχομένου ιστοσελίδων για αποδεικτικούς σκοπούς, δεν αναμένεται να αποτελέσει εξαίρεση σε αυτόν τον κανόνα, και για τον λόγο αυτό, στην παρούσα ενότητα θα κάνουμε διάφορες επισημάνσεις σχετικά με τεχνικά και άλλα θέματα που έχουν σχέση με αυτή καθέ αυτή την απόκρυξη δεδομένων (anti-forensics).

Πρώτα απ' όλα, με δεδομένο ότι το σύστημα ProCAVE θα τρέχει σε κεντρικούς εξυπηρετητές, κάποιος κακόβουλος μπορεί να βρει όλα τα ονόματα χώρου (domain names) και τις ηλεκτρονικές διευθύνσεις (ip) που το σύστημα αυτό χρησιμοποιεί, και να κατασκευάσει ιστοσελίδες που είτε απαγορεύουν την πρόσβαση του σε αυτές, είτε - ακόμα χειρότερα - παρουσιάζουν διαφορετικό (νομότυπο) περιεχόμενο όταν η επίσκεψη προέρχεται από το ProCAVE. Μία προφανής λύση σε αυτό θα ήταν η χρήση, από μεριάς μας, δυναμικών ηλεκτρονικών διευθύνσεων, ή άλλων ενδιάμεσων προγραμμάτων, που θα απέκρυπταν ή θα προκαλούσαν τη συνεχή αλλαγή των ηλεκτρονικών μας διευθύνσεων.

Μία άλλη τεχνική που θα μπορούσε να χρησιμοποιηθεί από κάποιον κακόβουλο θα ήταν η χρήση προγραμματιστικών εργαλείων και τεχνικών προκειμένου να παρουσιάσει στον χρήστη περιεχόμενο το οποίο δεν θα ήταν (τεχνικά) δυνατόν να αποθηκευτεί από το ProCAVE. Ένα τέτοιο παράδειγμα είναι η προβολή περιεχομένου μέσω βίντεο, η χρήση ασύγχρονου κώδικα κλπ. Παρ' όλα αυτά, πιστεύουμε ότι μία τέτοια κατάσταση είναι εύκολο να ξεπεραστεί, με μικρές τροποποιήσεις στην υλοποίηση του ProCAVE.

Για παράδειγμα, στην παρούσα υλοποίηση η εκτέλεση κώδικα ajax θα οδηγούσε στην τροποποίηση του περιεχομένου τοπικά, στον browser του χρήστη, χωρίς να γίνονται οι αντίστοιχες αλλαγές στο αποθηκευμένο στον Μηχανισμό Proxy περιεχόμενο. Μία λύση σε αυτό θα ήταν η υποστήριξη αναμονής γεγονότων (events listening) στον browser, και η αντίστοιχη εκτέλεση αυτών στον Μηχανισμό Proxy, με την χρήση τεχνικής που ονομάζεται scriptable web browser, όπως στο [25]. Ειδικά για την περίπτωση του κώδικα ajax, πιστεύουμε ότι είναι ήσσονος σημασίας, καθότι σύμφωνα με το [33], μόνο το 3.2% των ιστοσελίδων παγκοσμίως χρησιμοποιούν αυτήν την τεχνολογία.

Στο ίδιο πλαίσιο, υπάρχουν συγκεκριμένες εφαρμογές ή τεχνολογίες που δεν εμπίπτουν στην ανωτέρω κατηγορία· αυτές είναι κυρίως ιστοσελίδες που

χρησιμοποιούν teqnolog'ia flash, συγκεκριμένες εφαρμογές συνομιλιών chatting applications ή αποστολής μηνυμάτων instant messaging platforms κλπ. Σχετικά με αυτές, πιστεύουμε ότι είναι εκτός πεδίου εφαρμογής του συστήματος ProCAVE, καθώς εμπεριέχουν εικόνες ή/και κείμενο που εμφανίζεται στιγμιαία, και όχι το κλασικό web περιεχόμενο που υπάρχει στις ιστοσελίδες με τη μορφή που ξέρουμε.

Επίσης, από τη στιγμή που η λειτουργικότητα της πλατφόρμας ProCAVE θα είναι ελεύθερα διαθέσιμη στο κοινό, θα είναι έκθετη σε κάθε είδος ηλεκτρονικής επίθεσης, όπως αυτή της άρνησης υπηρεσιών (denial-of-service) κλπ. Για την αποφυγή τέτοιων περιστατικών, πρέπει να εφαρμοστούν κλασικές μέθοδοι προστασίας της ιστοσελίδας, όπως firewalls, intrusion prevention systems κλπ. Επιπλέον, η πρόσβαση στην πλατφόρμα ProCAVE μπορεί να περιοριστεί μόνο στους έχοντες κάποιο αποδεικτικό ηλεκτρονικής «ταυτοπροσωπίας», όπως ψηφιακό πιστοποιητικό, οι οποίοι θα μπορούν να πραγματοποιούν μόνο ένα ορισμένο αριθμό αιτήσεων ανά μονάδα χρόνου.

Επιπρόσθετα, και σχετικά με τη χρήση των ψηφιακών πιστοποιητικών για τη λειτουργία της επιλογής ιδιωτικότητας του συστήματος ProCAVE, είναι φανερό ότι πρέπει να υπάρξει μία Υποδομή Δημοσίου Κλειδιού (Public-Key Infrastructure - PKI), για τη δημιουργία, διαχείριση και απόσυρση των πιστοποιητικών αυτών. Παρ' όλα αυτά, το επίπεδο το εμπιστοσύνης που θα επιτύχουμε εξαρτάται από την επιλογή της Αρχής Πιστοποίησης (Certificate Authority - CA) που θα κάνουμε. Σε απλές εγκαταστάσεις της πλαρφόρμας ProCAVE, μπορούν να χρησιμοποιηθούν εργαλεία ανοικτού κώδικα, όπως το OpenSSL ([23]), αλλά σε εγκαταστάσεις μεγαλύτερης κλίμακας χρειάζεται μία κοινά αποδεκτή οντότητα πιστοποίησης.

Τέλος, κάποιος θα μπορούσε να αναρωτηθεί σχετικά με την αρχή ή τον οργανισμό που θα μπορούσε να θεωρηθεί αξιόπιστος ώστε να «φιλοξενεί» το σύστημα ProCAVE και να κρατά τοπικά αντίγραφα του περιεχομένου ιστοσελίδων. Πιστεύουμε ότι το θέμα αυτό πρέπει να χειριστεί ανάλογα με τη νομοθεσία κάθε χώρας. Μία ενδεδειγμένη λύση θα ήταν η χρήση του ProCAVE από τα επίσημα εγκληματολογικά εργαστήρια (forensics labs) κάθε χώρας. Παρ' όλα αυτά, ο τρόπος χειρισμού των πειστηρίων από το ProCAVE, το γεγονός ότι μπορεί να τρέξει παράλληλα σε πολλές γεωγραφικές τοποθεσίες καθώς και η επιλογή ιδιωτικότητας που διαθέτει, μας κάνουν να συμπεράνουμε ότι όσο περισσότερες εγκαταστάσεις του ProCAVE υπάρξουν, τόσο καλύτερα για την αξιοπιστία και τη διαθεσιμότητα των πειστηρίων.

Κεφάλαιο 5

Επίλογος

Στο παρόν κεφάλαιο θα κάνουμε μία ανακεφαλαίωση της εργασίας, και θα καταγράψουμε τα συμπεράσματα στα οποία μας οδήγησε η μελέτη των εργαλείων και των μεθόδων συλλογής και πιστοποίησης του περιεχομένου ιστοσελίδων. Κλείνοντας, θα αναφερθούμε στα επόμενα βήματα έρευνας και ανάπτυξης στον συγκεκριμένο τομέα.

5.1 Συμπεράσματα

Στην παρούσα εργασία κάναμε μία εκτενή αναφορά στο ζήτημα της αποθήκευσης και πιστοποίησης του περιεχομένου ιστοσελίδων για αποδεικτικούς σκοπούς, δηλαδή στην συλλογή και διαχείριση του περιεχομένου κατά τέτοιο τρόπο ώστε να αποδεικνύει, εκτός από τη μορφή του ίδιου του περιεχομένου, το πότε αυτό αποθηκεύτηκε, από ποιον, με ποια μέθοδο, αλλά και να διατηρεί ταυτόχρονα και την ιδιωτικότητα του χρήστη.

Αρχικά κάναμε μία εκτενή αναφορά στην τρέχουσα βιβλιογραφία, και στα προβλήματα αποδεικτικότητας που υπάρχουν εγγενώς σε όλα τα πειστήρια που προέρχονται από ιστοσελίδες του διαδικτύου, λόγω της δυναμικής και συνεχώς μεταλλασσόμενης φύσης του. Στο πλαίσιο αυτό, εντοπίσαμε και καταγράψαμε τις ιδιότητες που πρέπει να πληρούν οι μέθοδοι συλλογής και πιστοποίησης περιεχομένου, προκειμένου να παρέχουν πειστήρια πέρα από κάθε αμφισβήτηση.

Εν συνεχεία αναφερθήκαμε στις πιο συχνά χρησιμοποιούμενες μεθόδους μέχρι σήμερα, που είναι η τοπική αποθήκευση, η μαγνητοσκόπηση πλοήγησης, η χρήση τρίτων υπηρεσιών, και η χρήση ιστοσελίδων διατήρησης περιεχομένου. Για κάθε μία από αυτές, αναλύσαμε τα χαρακτηριστικά τους και επιχειρηματο-

λογήσαμε για το κατά πόσο πληρούν τα κριτήρια τα οποία θέσαμε. Συμπεράναμε ότι όλες υπολείπονται στην πλήρωση ενός ή περισσοτέρων κριτηρίων, αλλά ειδικά το κριτήριο της διαφύλαξης της ιδιωτικότητας του χρήστη απουσιάζει παντελώς από όλες.

Ακολούθως παρουσιάσαμε μία νέα μέθοδο συλλογής και διατήρησης περιεχομένου ιστοσελίδων που προτείνουμε, με την ονομασία ProCAVE. Αναφερθήκαμε στα κύρια χαρακτηριστικά λειτουργίας της μεθόδου αυτής, και σταθήκαμε ιδιαίτερα στην αμεσότητα στη χρήση που παρέχει, στην χρονοσήμανση και αυθεντικότητα, και - κυρίως - στη διαφύλαξη της ιδιωτικότητας του χρήστη, πληρώντας έτσι όλα τα τεθέντα κριτήρια.

Τέλος, κάναμε αναφορά σε μία πρότυπη υλοποίηση ενός συστήματος που πραγματοποιήσαμε, το οποίο είναι βασισμένο στη μέθοδο ProCAVE που προτείνουμε. Με τη χρήση του συστήματος αυτού, διεξάγαμε μία σειρά πειραμάτων συλλογής και πιστοποίησης γνωστών ιστοσελίδων, με σκοπό να αξιολογήσουμε την ορθότητα των ισχυρισμών μας. Όπως προκύπτει από τα αποτελέσματα των πειραμάτων, τα συλλεχθέντα περιεχόμενα παρουσιάζουνε αποδεδειγμένα μεγάλη ομοιότητα με αυτά που παρουσιάζονται στον χρήστη.

Βάσει αυτού συμπεραίνουμε ότι η μέθοδος ProCAVE μπορεί να παράξει αδιαμφηβήτητα πειστήρια σε πραγματικό χρόνο, την ώρα δηλαδή που ο χρήστης πλοηγείται στο διαδίκτυο και διαπιστώνει κάποιο περιεχόμενο που μπορεί να χρησιμοποιηθεί σε μία αποδεικτική διαδικασία, και το σημαντικότερο, χωρίς να εμπλέκονται πολύπλοκοι μηχανισμοί και λειτουργίες. Γίνεται μνεία ότι μέρος της παρούσας εργασίας, ως επιστημονικό άρθρο έγινε αποδεκτό για δημοσίευση στα πρακτικά του 10th International Conference on Trust, Privacy & Security in Digital Business (TrustBus 2013) [17].

5.2 Επόμενα βήματα

Τα επόμενα βήματα της εργασίας μας θα αφορούν την επέκταση του συστήματος που αναπτύξαμε προκειμένου να συμπεριλάβει λειτουργίες που αυτή τη στιγμή δεν υποστηρίζονται, όπως η συλλογή περιεχομένου ιστοσελίδων με βίντεο, παιχνίδια κλπ. Επίσης, θεωρούμε ως σημαντικό επόμενο στάδιο την υιοθέτηση και χρήση της λύσης μας από χρήστες του διαδικτύου, προκειμένου να αναγνωριστούν πιθανές ελλείψεις ή σημεία προς βελτίωση.-

Παράρτημα Α΄

Επιλεγμένος Κώδικας

Κώδικας 1 : Υλοποίηση του Μηχανισμού Proxy

```
1  function PIPHP_SimpleWebProxy($url, $redirect)
2  {
3  global $ses;
4
5  $ch = curl_init();
6
7  $contents = @file_get_contents($url);
8  if (!$contents) return NULL;
9
10 $purl = parse_url($url);
11
12 switch(strtolower(substr($url, -4)))
13 {
14 case ".jpg": case ".gif": case ".png": case ".ico":
15 case ".js": case ".xml":
16 return $contents;
17 case ".css":
18 return $contents =
19 str_replace("..", $redirect."&u=".$purl['scheme'].
20 "://".$purl['host'], $contents);
21 }
22 $ctofile = $contents;
```

```
25 26
23 $contents = str_replace('_blank',     '!!**2**!!', $contents);
24 $contents = str_replace('&amp;', '&',         $contents);
27 $contents = str_replace('&',     '!!**1**!!', $contents);
28
29 $dom      = new domdocument();
30 @$dom     ->loadhtml($contents);
31 $xpath    = new domxpath($dom);
32 $hrefs    = $xpath->evaluate("/html/body//a");
33 $sources  = $xpath->evaluate("/html/body//img");
34 $iframes  = $xpath->evaluate("/html/body//iframe");
35 $scripts  = $xpath->evaluate("/html//script");
36 $css      = $xpath->evaluate("/html/head//link");
37 $links    = array();
38
39 for ($j = 0 ; $j < $hrefs->length ; ++$j)
40 $links[] = $hrefs->item($j)->getAttribute('href');
41
42 for ($j = 0 ; $j < $sources->length ; ++$j)
43 $links[] = $sources->item($j)->getAttribute('src');
44
45 for ($j = 0 ; $j < $iframes->length ; ++$j)
46 $links[] = $iframes->item($j)->getAttribute('src');
47
48 for ($j = 0 ; $j < $scripts->length ; ++$j)
49 $links[] = $scripts->item($j)->getAttribute('src');
50
51 for ($j = 0 ; $j < $css->length ; ++$j)
52 $links[] = $css->item($j)->getAttribute('href');
53
54
55 $links = array_unique($links);
56 $to    = array();
57 $count = 0;
58 sort($links);
59
60 foreach ($links as $link)
61 {
```

```
63 64
62 if ($link != "")
65 {
66 $temp = str_replace('!!**1**!!', '&', $link);
67 $ll = basename($link);
68 $link = base64_encode($url);
69
70 $to[$count] = "/$redirect&u=" .
71   urlencode(PIPHP_RelToAbsURL($url, $temp));
72 $contents = str_replace("href=\"$link\"",
73    "href=\"!!$count!!\"", $contents);
74 $contents = str_replace("href='$link'",
75    "href='!!$count!!'",   $contents);
76 $contents = str_replace("href=$link",
77    "href=!!$count!!",     $contents);
78 $contents = str_replace("src=\"$link\"",
79    "src=\"!!$count!!\"",  $contents);
80 $contents = str_replace("src='$link'",
81    "src='!!$count!!'",    $contents);
82 $contents = str_replace("src=$link",
83    "src=!!$count!!",      $contents);
84 ++$count;
85
86 $ctofile = str_replace("href=\"$link\"",
87    "href=\"$ll\"", $ctofile);
88 $ctofile = str_replace("href='$link'",
89    "href='$ll'",   $ctofile);
90 $ctofile = str_replace("href=$link",
91    "href=$ll",     $ctofile);
92 $ctofile = str_replace("src=\"$link\"",
93    "src=\"$ll\"",  $ctofile);
94 $ctofile = str_replace("src='$link'",
95    "src='$ll'",    $ctofile);
96 $ctofile = str_replace("src=$link",
97    "src=$ll",      $ctofile);
98 }
99 }
100
```

```
102 103
101 $ext = strtolower(substr(basename($url), -4));
104 if( $ext == "html" || $ext == ".htm"  )
105 PIPHP_DumpToFile( basename($url), $ctofile, $ses);
106 else
107 PIPHP_DumpToFile( basename($url).".html", $ctofile, $ses);
108
109 for ($j = 0 ; $j < $count ; ++$j)
110 $contents = str_replace("!!$j!!", $to[$j],
111    $contents);
112
113 return str_replace('!!**1**!!', '&', $contents);
114 }
```

Κώδικας 2 : Υλοποίηση του Μηχανισμού CVA

```
1  function PIPHP_DumpToFile($filename, $content, $hex){
2
3  $path = "store/";
4  $now = time();
5  $hash1 = $hash2 = "";
6  $signature = "";
7
8  if(!file_exists($path.$hex))
9  mkdir($path.$hex);
10
11 // Create file
12 $fp = fopen($path.$hex."/".$filename, 'w');
13 fwrite($fp, $content);
14 fclose($fp);
15
16 // Create timestamp
17 $fp = fopen($path.$hex."/".$filename.".time", 'w');
18 fwrite($fp, $now);
19 fclose($fp);
20
21 // Create hash1 of the file
22 $fp = fopen($path.$hex."/".$filename.".ripemd320_1", 'w');
23 $hash1 = hash_file('ripemd320',$path.$hex."/".$filename);
24 fwrite($fp,$hash1);
25 fclose($fp);
26
27 // Create hash1+time of the file
28 $fp = fopen($path.$hex."/".$filename.".ripemd320_1plustime", 'w');
29 fwrite($fp,$hash1.$now);
30 fclose($fp);
31
32 // Create hash2 of hash1+time
33 $fp = fopen($path.$hex."/".$filename.".ripemd320_2", 'w');
34 $hash2 = hash('ripemd320',$hash1.$now);
35 fwrite($fp,$hash2);
36 fclose($fp);
```

```
38 39
37
40 // Sign hash2 with time server private key
41 $fp = fopen("Time_priv.pem", "r");
42 $priv_key = fread($fp, 8192);
43 fclose($fp);
44 $pkeyid = openssl_get_privatekey($priv_key);
45 if(empty($pkeyid)){
46 die("Can't load key id");
47 }
48 if(!openssl_sign($hash2, $signature, $pkeyid, OPENSSL_ALGO_SHA1) ){
49 die("Failed to sign data: $hash2");
50 }
51 openssl_free_key($pkeyid);
52 if(empty($signature)){
53 die("signature empty");
54 }
55 // Create hash2 of hash1+time
56 $fp = fopen($path.$hex."/".$filename.".final", 'w');
57 fwrite($fp,$signature);
58 fclose($fp);
59
60 }
```

Βιβλιογραφία

[1] Berghel, H. Hiding data, forensics, and anti-forensics. *Communications of the ACM 50*, 4 (Apr. 2007), 15–20.

[2] Canprove - capture online evidence. `http://canprove.com`. [Online. Accessed: Sep. 2013].

[3] Careless, J. Collecting and authenticating online evidence (cba practicelink). `http://www.cba.org/cba/practicelink/careerbuilders_technology/web-evidence.aspx`. [Online. Accessed: Sep. 2013].

[4] Casey, E. *Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet*, 3rd ed. Academic Press, 2011.

[5] Commonwealth vs. Michael M. O'Laughlin: Burglary, armed assault in a dwelling, assault and battery by means of a dangerous weapon, practice, criminal, required finding, 2005. Appellate Court Decision, No. 04-P-48.

[6] Council of Europe (CoE), Explanatory Report to the Convention on Cybercrime (ETS 185). `http://conventions.coe.int/Treaty/en/Reports/Html/185.htm`, 2001. [Online. Accessed: Sep. 2013].

[7] Dvers - digital verification services. `http://www.dvers.gr`. [Online. Accessed: Sep. 2013].

[8] Fenner, M. Evidentiary problems associated with the introduction of web-based evidence (lsn: Evidence (public law) sub-topic). `http://ssrn.com/abstract=1722714`, December 2010. [Online. Accessed: Sep. 2013].

[9] U.S.Courts - Federal Rules of Evidence. http://www.uscourts.gov/uscourts/RulesAndPolicies/rules/2010\%20Rules/Evidence.pdf, December 2010. [Online. Accessed: Sep. 2013].

[10] GARFINKEL, S. L. Digital forensics research: The next 10 years. *Digital Investigations 7* (Aug. 2010), S64–S73.

[11] GIBSON, J. A primer on admitting web page into evidence (nevada lawyer magazine). http://nvbar.org/articles/content/primer-admitting-web-pages-evidence. [Online. Accessed: Sep. 2013].

[12] Hanzo Archives. http://www.hanzoarchives.com. [Online. Accessed: Sep. 2013].

[13] HOSMER, C. Proving the integrity of digital evidence with time. *IJDE 1*, 1 (2002).

[14] KANELLIS, P., KIOUNTOUZIS, E., KOLOKOTRONIS, N., AND MARTAKOS, D. *Digital Crime And Forensic Science in Cyberspace*. IGI Publishing, Hershey, PA, USA, 2006.

[15] KARYDA, M., AND MITROU, L. Internet forensics: Legal and technical issues. In *Digital Forensics and Incident Analysis (WDFIA), 2007* (2007), pp. 3–12.

[16] KERZNER, M. Evidence authentication: Web site content (atkison baker). http://ssrn.com/abstract=1722714. [Online. Accessed: Sep. 2013].

[17] LALAS, E., MITROU, L., AND LAMBRINOUDAKIS, C. Procave: Privacy-preserving collection and authenticity validation of online evidence. In *Trust, Privacy, and Security in Digital Business*, S. Furnell, C. Lambrinoudakis, and J. Lopez, Eds., vol. 8058 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2013, pp. 137–148.

[18] LEROUX, O. Legal admissibility of electronic evidence1. *International Review of Law, Computers and Technology 18*, 2 (2004), 193–220.

[19] Jack R. Lorraine and Beverly Mack, Plaintiffs vs. Markel American Insurance Company, Defendants. Civil Action No. PWG-06-1893, 2007. United States District Court for the District of Maryland.

[20] Mylonas, A., Meletiadis, V., Mitrou, L., and Gritzalis, D. Dynamic evidence acquisition for smartphone forensics. In *27th IFIP International Information Security and Privacy Conference, Springer (AICT 267)* (2012), pp. 245–256.

[21] Mylonas, A., Meletiadis, V., Mitrou, L., and Gritzalis, D. Smartphone sensor data as digital evidence. *Computers and Security 38*, 0 (2013), 51–75. Cybercrime in the Digital Economy.

[22] Nixon, R. *Plug-In PHP: 100 Power Solutions : Simple Solutions to Practical PHP Problems*. Mcgraw-hill, 2010.

[23] Openssl - the cryptography and ssl/tls toolkit. `http://www.openssl.org`. [Online. Accessed: Sep. 2013].

[24] PageFreezer. `http://pagefreezer.com`. [Online. Accessed: Sep. 2013].

[25] Php scriptable web browser. `http://www.simpletest.org/en/browser_documentation.html`. [Online. Accessed: Sep. 2013].

[26] ProCAVE. `http://www.ds.unipi.gr`. [Available upon request].

[27] Reed Archives. `http://www.reedarchives.com`. [Online. Accessed: Sep. 2013].

[28] Similarity analyzer. `http://tool.motoricerca.info/similarity-analyzer.phtml`. [Online. Accessed: Sep. 2013].

[29] Sommer, P. *Digital Evidence, Digital Investigations and E-Disclosure: A Guide to Forensic Readiness for Organisations, Security Adviders and Lawyers, Third Edition, Version 3.0*. Information Assurance Advisory Council, 2012.

[30] Tanenbaum, A. *Computer Networks*, 4th ed. Prentice Hall Professional Technical Reference, 2002.

[31] The internet archive. `http://archive.org`. [Online. Accessed: Sep. 2013].

[32] Using bb flashback in cyber forensics. `http://www.bbsoftware.co.uk/Solutions/CyberForensics.aspx`. [Online. Accessed: Sep. 2013].

[33] Web statistics. http://www.scriptol.com/web/statistics.php. [Online. Accessed: Sep. 2013].

[34] X1 social discovery. http://www.x1discovery.com/social_discovery.html. [Online. Accessed: Sep. 2013].