# CVE-2016-5195



## הסבר כללי

חולשה שהייתה קיימת בקרנל של לינוקס מאז 2007. החולשה מאפשרת כתיבה לקבצים שבהרשאות קריאה-בלבד. בכלליות החולשה מנצלת race condition שקיים בצורת המימוש של מנגנון הCOW בקרנל.

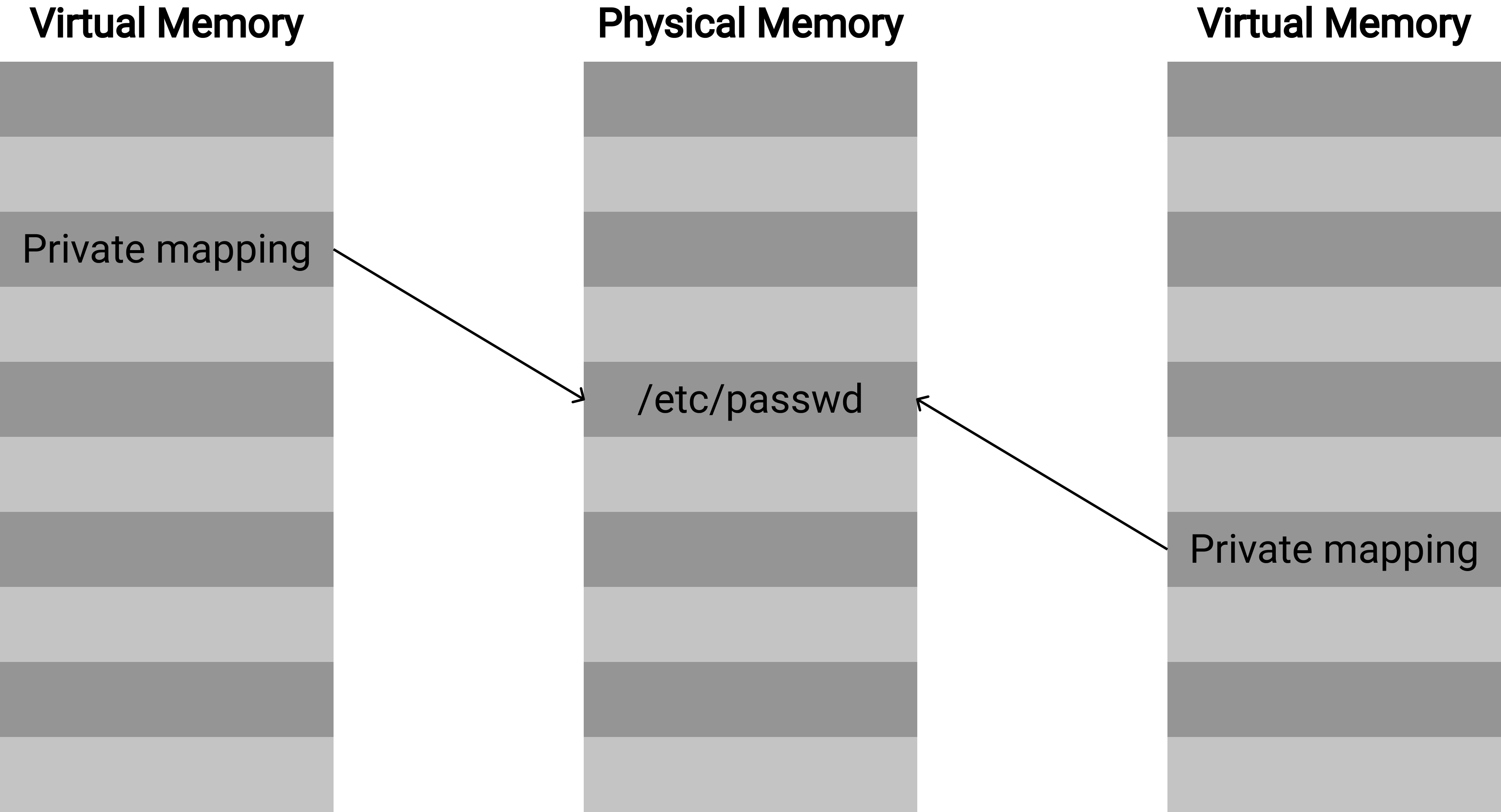
## מושגים בסיסיים

### מרוץ תהליכים

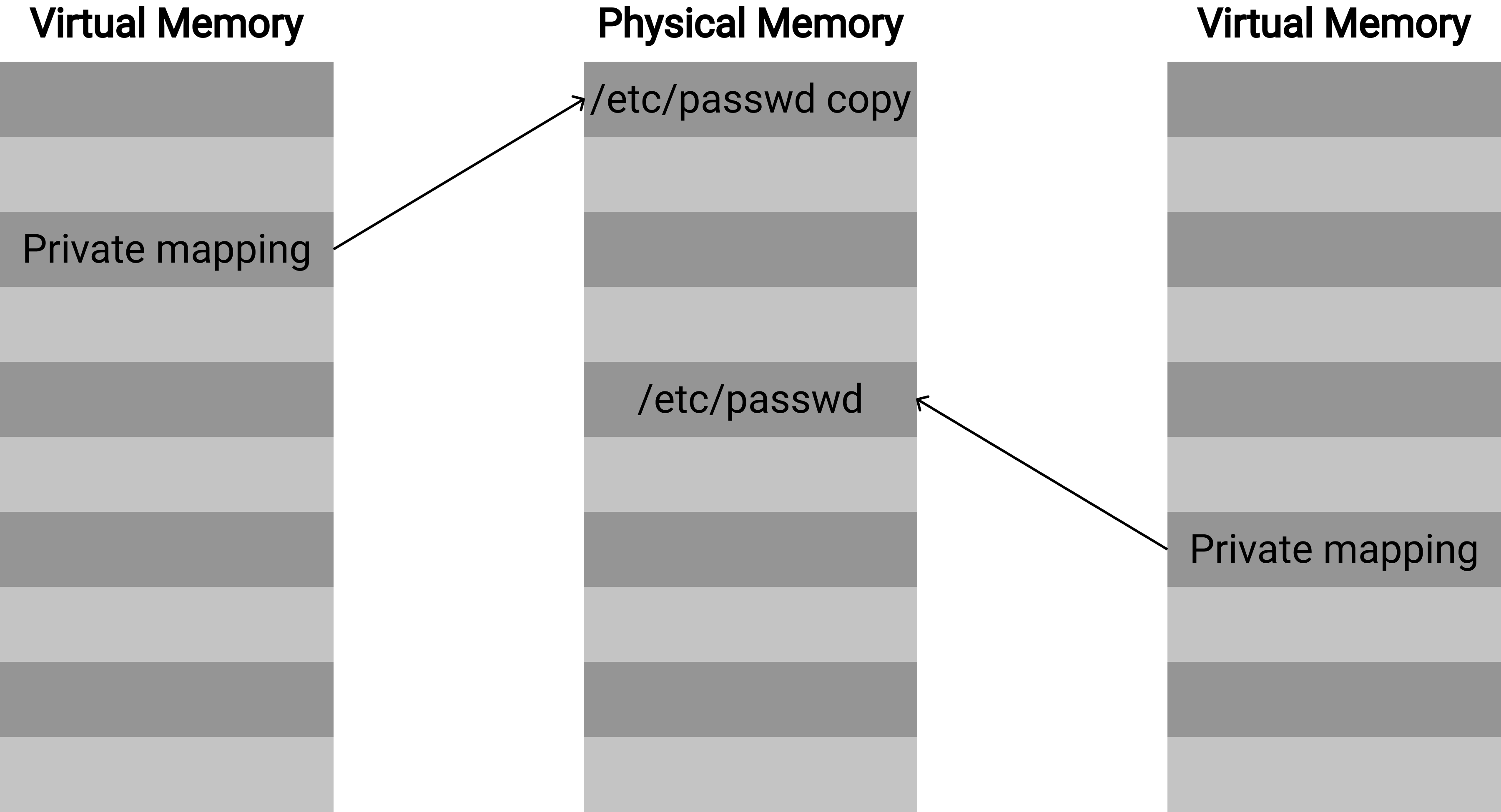
ב[מדעי המחשב](https://he.wikipedia.org/wiki/%D7%9E%D7%93%D7%A2%D7%99_%D7%94%D7%9E%D7%97%D7%A9%D7%91), **מרוץ תהליכים** (Race condition) או לעיתים **תחרות** או **תנאי מרוץ**, הוא מצב שבו מספר [תהליכים](https://he.wikipedia.org/wiki/%D7%AA%D7%94%D7%9C%D7%99%D7%9A_(%D7%9E%D7%93%D7%A2%D7%99_%D7%94%D7%9E%D7%97%D7%A9%D7%91)) או [תהליכונים](https://he.wikipedia.org/wiki/%D7%AA%D7%94%D7%9C%D7%99%D7%9B%D7%95%D7%9F) מנסים בו זמנית להשיג גישה למשאב משותף. כיוון שהגישה למשאב המשותף נעשית במסגרת [קטע קריטי](https://he.wikipedia.org/wiki/%D7%A7%D7%98%D7%A2_%D7%A7%D7%A8%D7%99%D7%98%D7%99), המאפשר גישה רק למספר מוגבל של תהליכים או תהליכונים (ולרוב רק לאחד), במצב כזה מתקיימת תחרות בתנאי כל הקודם זוכה.

### העתקה בזמן כתיבה

נגיד ויש שני תהליכים שניגשים לאותו המשאב שטעון לזיכרון הפיזי. במידה ושניהם רוצים רק לקרוא מהמשאב, אין שום סיבה לטעון אותו פעמיים, לכן עבור כל תהליך תהיה כתובת וירטואלית שממופה לאותה הכתובת בזיכרון. על מנת לאפשר COW ישלח גם הדגל MAP\_PRIVATE בעת מיפוי המשאב. כמו שניתן לראות, הכתובת הווירטואלית הממופה של תהליך #1 והכתובת הווירטואלית הממופה של תהליך #2 מצביעות אל אותו הדף בזיכרון.



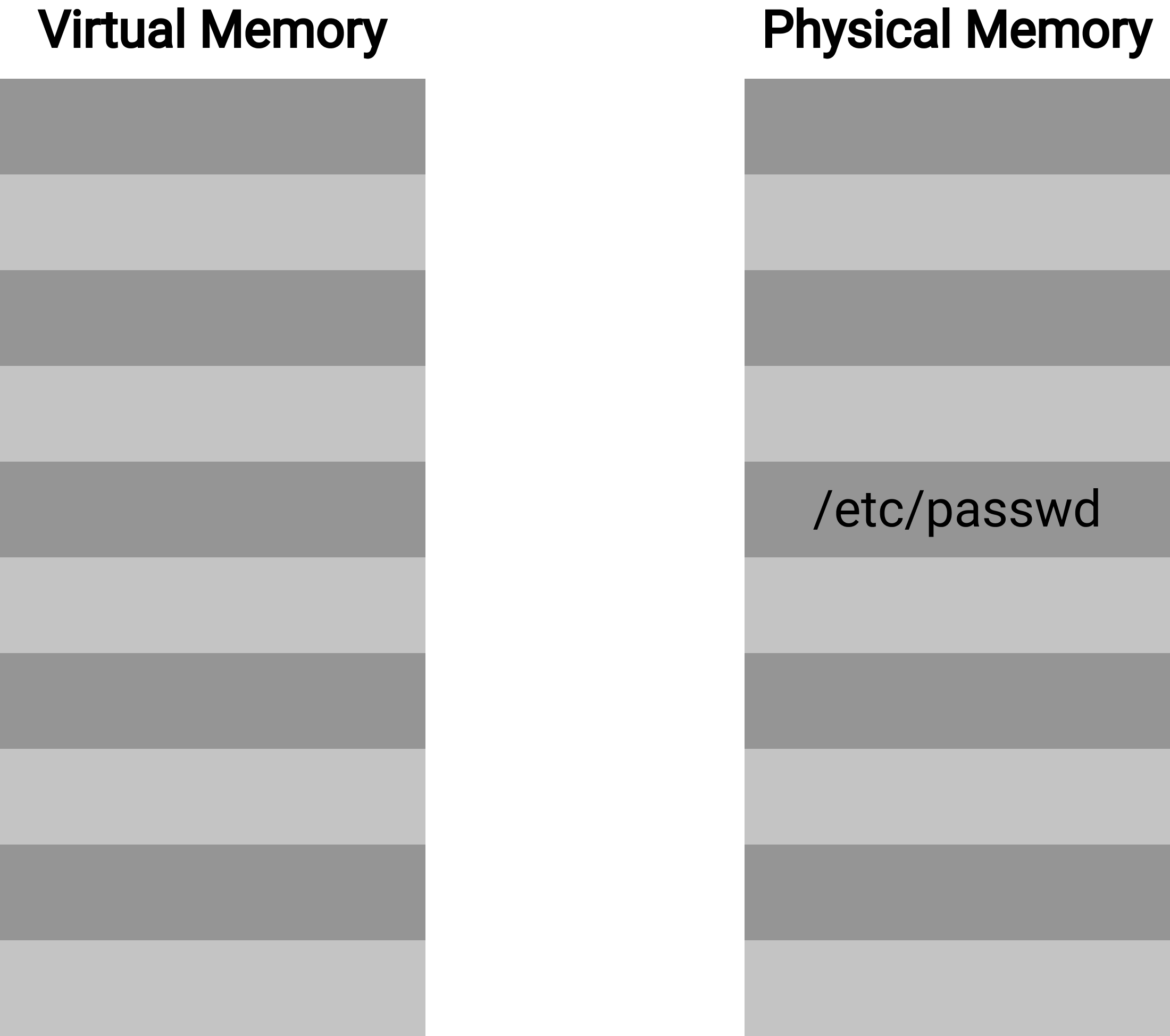
מה יקרה אבל כאשר תהליך אחד ירצה לשנות את אותו המשאב שממופה לו לזיכרון? סביר לגמרי שכמתכנת, לא ארצה לאפשר לתהליך אחד להתעסק עם משאב שתהליך אחר מסתמך עליו. במקרה כזה ייווצר עותק חדש של המשאב בזיכרון הפיזי, והמיפוי בזיכרון הווירטואלי של אותו התהליך יצביע אל העותק החדש.



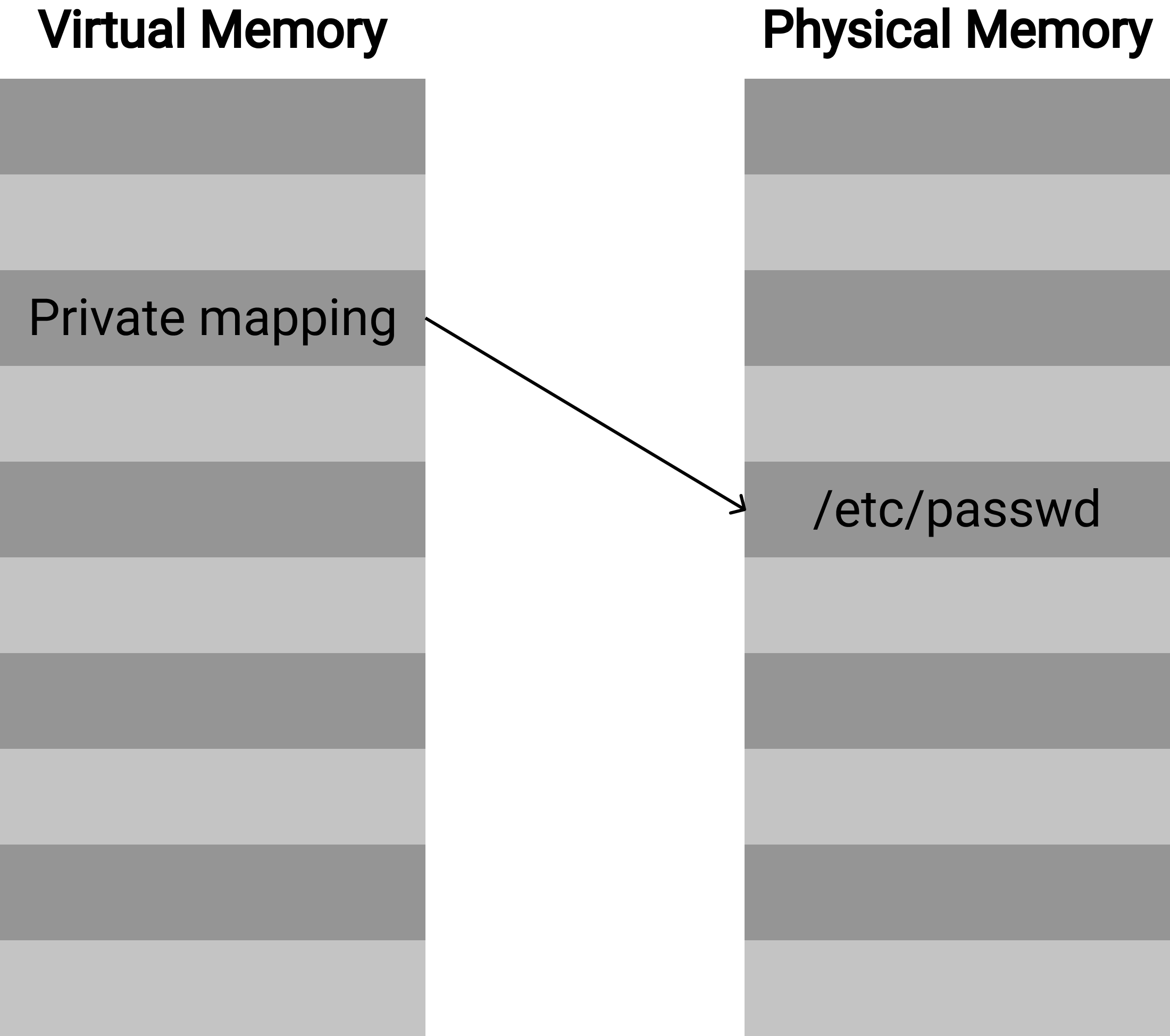
חשוב להדגיש שבדרך כלל, כל עותק שנוצר בעקבות COW הוא פרטי עבור תהליך, משמע שכל שינוי שיתבצע עליו לא יחלחל למשאב המקורי.

## כיצד ממומש הExploit

ראשית בקוד נפתח הקובץ בהרשאות קריאה בלבד ונטען לזיכרון. ניקח לדוגמה את הקובץ /etc/passwd. כידוע אנחנו כמשתמשים פשוטים לא יכולים לשנות אותו.



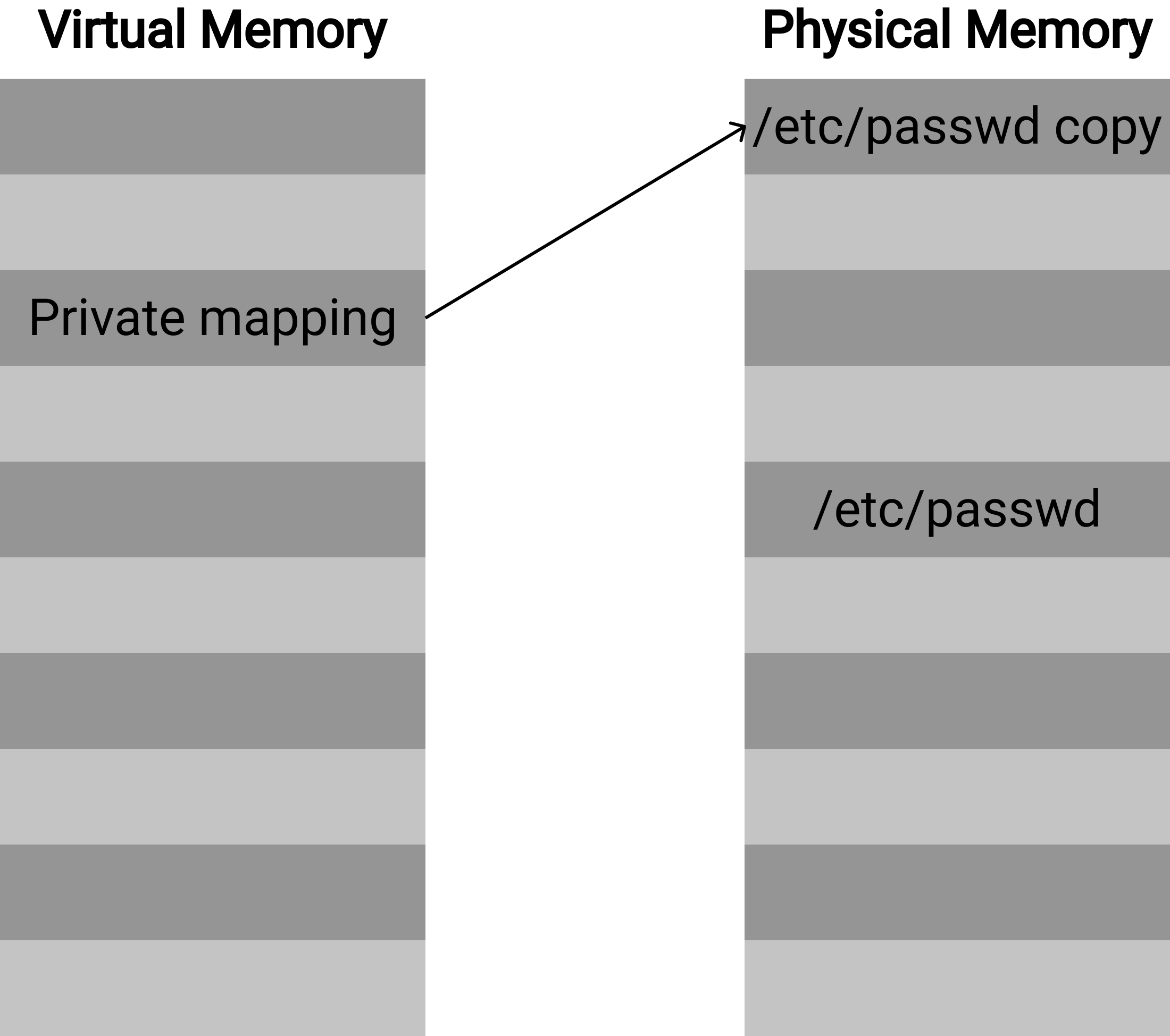
כעת התהליך יבקש ליצור מיפוי פרטי של אותו הקובץ למרחב הכתובות הווירטואלי שלו בעזרת הפונקציה mmap. בפונקציה יועבר גם הדגל MAP\_PRIVATE על מנת שיוכל להשתמש בCOW. מכיוון שיצרתי את המיפוי עם אותו הדגל, נוכל לכתוב שינויים להעתק פרטי של אותו הקובץ למרות שהוא בעיקרון לקריאה בלבד.



בשלב הזה נכנס הrace condition. הקוד מכניס שני תהליכונים שונים ללולאות בהן אחד מנסה לכתוב למיפוי הקובץ, ותהליכון שני מריץ את הSyscall 'madvice' עם הדגל MADV\_DONTNEED. הsyscall הזה מציע לקרנל מה עליו לעשות עם דפים מסויימים, והדגל הזה אומר את זה:

Do not expect access in the near future. (For the time being, the application is finished with the given range, so the kernel can free resources associated with it.)

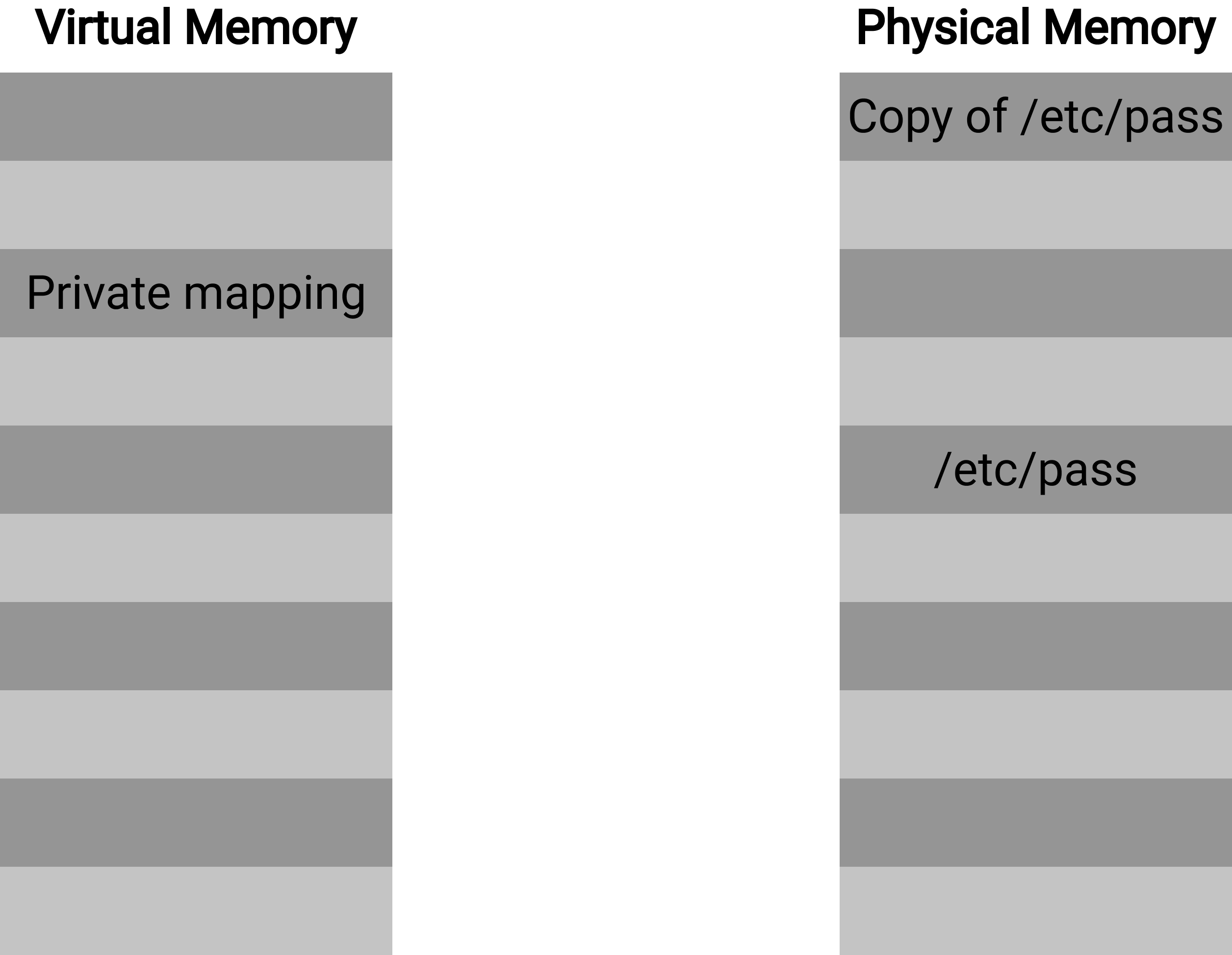
כאשר תהליך ינסה לכתוב למיפוי הפרטי הזה, הקרנל קודם כל ייצור עותק של אותו הדף בזיכרון הפיזי, ישנה את מיפוי הכתובות בזיכרון הווירטואלי ככה שיצביע על הדף החדש (המועתק), ואז יכתוב לתוכו. אך מרוץ התהליכים מצליח לעקוף את השלב של יצירת העותק של הקובץ הזיכרון ולכתוב ישירות לקובץ המקורי.



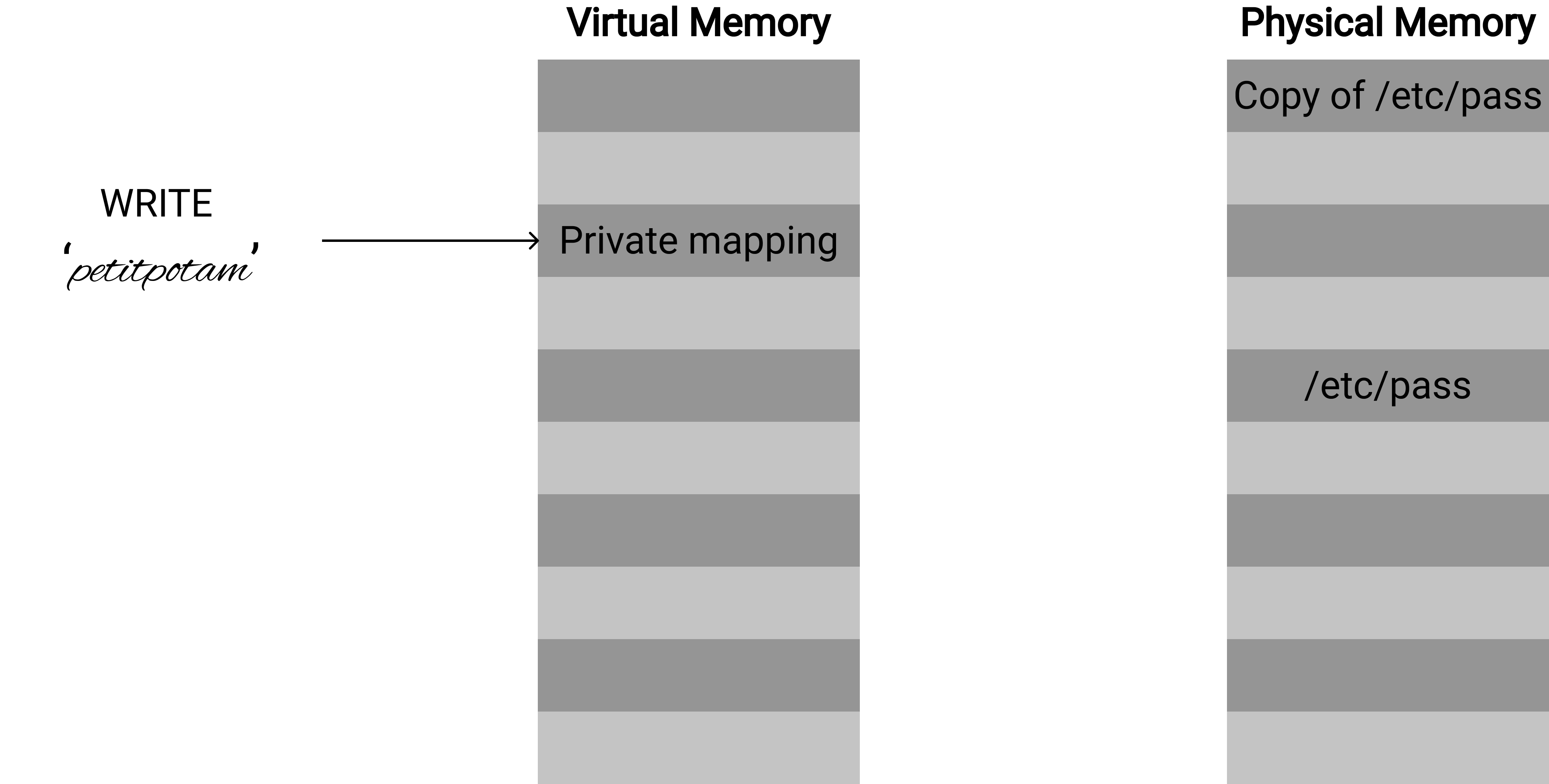
### מרוץ התהליכים

התהליך עכשיו יוצר שני תהליכונים (Threads) שירוצו במקביל ויעשו סט פעולות מסוים אלפי פעמים על מנת לנסות ליצור סיטואציה שהסבירות שתקרה בדרך כלל מאוד נמוכה. זה הוא החלק בחולשה שהוא "מרוץ תהליכים (Race condition).

התהליכון הראשון מריץ את הSyscall 'madvice' ומציע לקרנל איך עליו להתנהג ביחס לדפים מסוימים בזיכרון הווירטואלי. ההצעה שנשלחת לקרנל היא MADV\_DONTNEED כלומר שאין יותר צורך באותו המיפוי לקובץ שנוצר בזיכרון הווירטואלי, ולכן אפשר לשכוח ממנו.



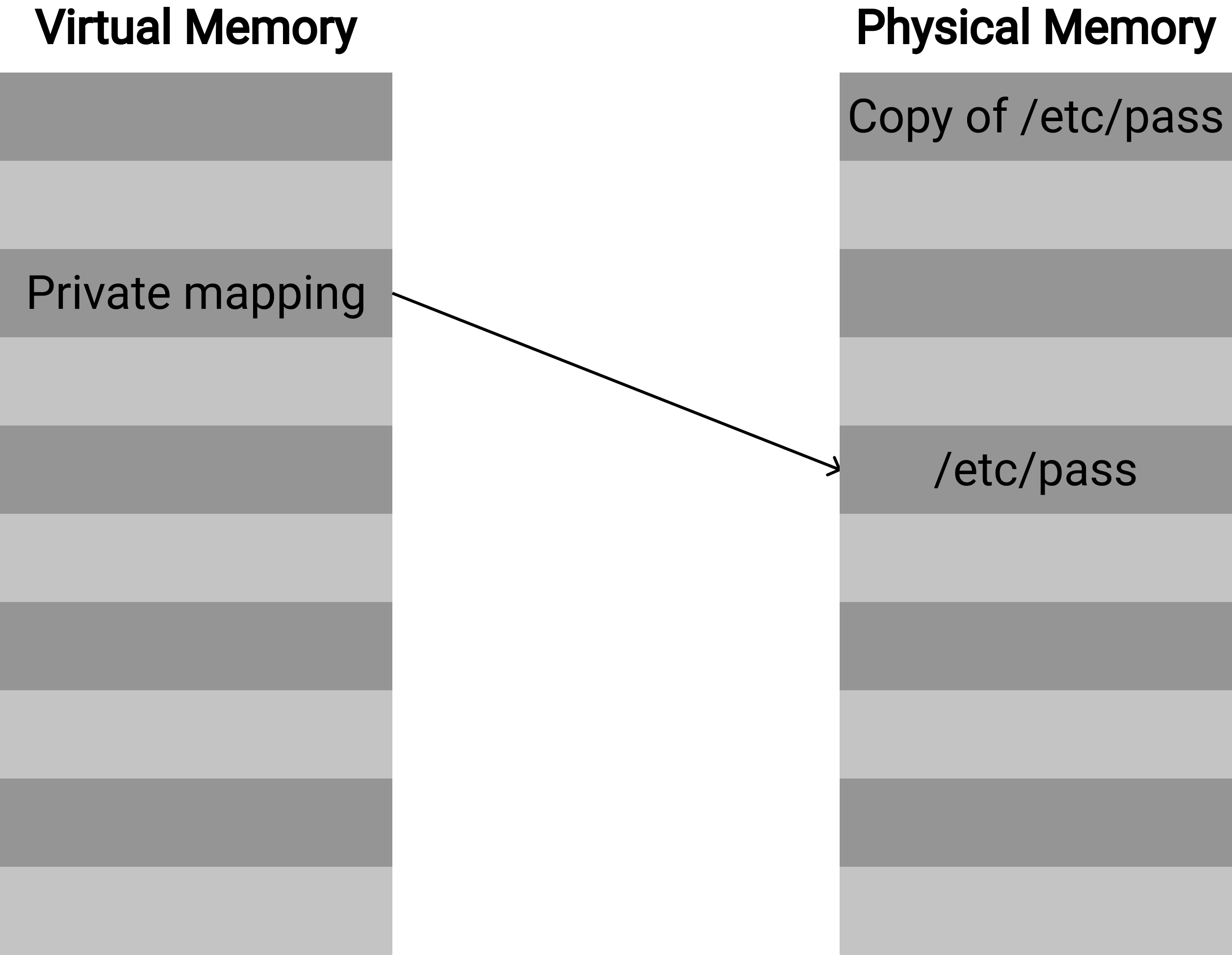
התהליכון השני פותח קובץ מיוחד במערכת ההפעלה /proc/self/mem שהוא בעצם ייצוג של הזיכרון הווירטואלי של התהליך הנוכחי במערכת הקבצים. לאחר מכן הוא מנסה לכתוב אל תוך המיפוי של הקובץ (/etc/pass).



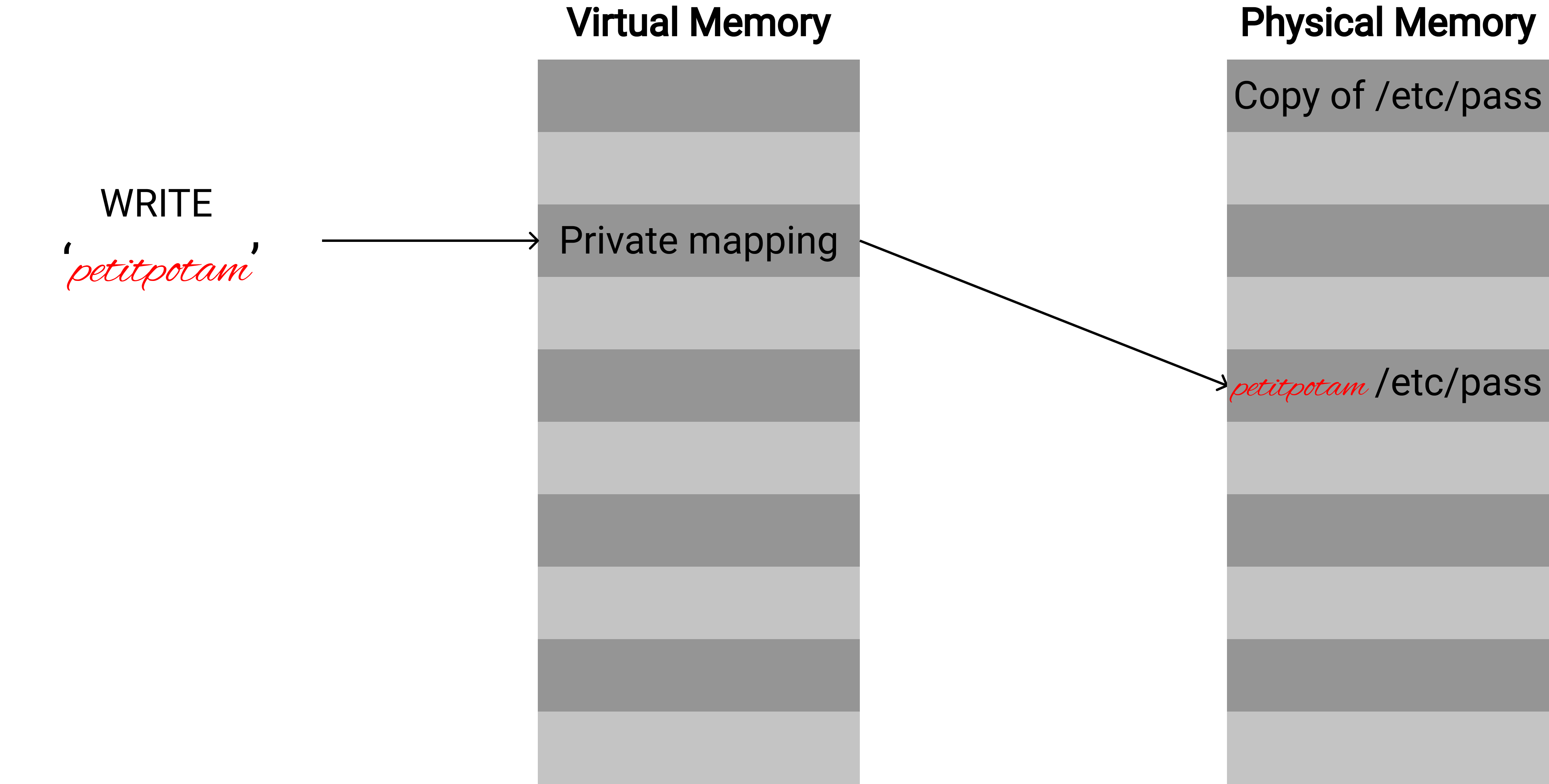
ההגדרה של MADV\_DONTNEED בman אומרת ש:

subsequent accesses of pages in the range will succeed, but will result in either repopulating the memory contents from the up-to date contents of the underlying mapped file.

כלומר לאחר שהקרנל נפטר מהמיפוי הפרטי של הדף בזיכרון, בפעם הבאה שתתבצע פנייה לאותם הדפים בזיכרון הווירטואלי, הקובץ המקורי ימופה מחדש ולא יצביע לעותק שנוצר קודם לכן.



מה שאומר שאומר שבעת הקריאה לכתיבה לכתובות הממופות בזיכרון הווירטואלי, אותם הדפים ימופו לקובץ המקורי במקום לעותק פרטי, והכתיבה תתבצע על הקובץ המקורי.



## מקורות

<https://www.digitalwhisper.co.il/files/Zines/0x4D/DW77-1-RaceCondition.pdf>

<https://man7.org/linux/man-pages/man2/madvise.2.html>

<https://www.youtube.com/watch?v=CQcgz43MEZg>

<https://www.cs.toronto.edu/~arnold/427/18s/427_18S/indepth/dirty-cow/index.html>